



International Copyright Issues and Artificial Intelligence

Transcript from Online Webinar on July 26, 2023

Chris Weston: Hello, everyone, and thank you for joining us today for the Copyright Office's webinar on International Copyright Issues Arising from Artificial Intelligence Technology. I'm Chris Weston, Senior Counsel in the Office of Policy and International Affairs. To kick off today's webinar, it is my pleasure to introduce Shira Perlmutter, Register of Copyrights and Director of the U.S. Copyright Office for opening remarks.

Shira Perlmutter: Thanks, Chris. Welcome to the Copyright Office's webinar on International Copyright Issues and Artificial Intelligence. Today's webinar is part of the Office's ongoing initiative to examine the Copyright Law and policy issues raised by AI technology. We've already hosted a series of public events examining these issues. We held four listening sessions in April and May, and heard from a wide range of stakeholders about their experiences with AI in connection with literary works, visual arts, audiovisual works and musical works and sound recordings. Last month, we hosted a webinar to walk the public through our March 16th registration guidance. And, at that time, we offered numerous examples to help applicants understand how to register works containing material produced by generative AI. If you're interested and missed any of these events, you can find recordings, as well as our registration guidance, on the Copyright Office website.

Now, while these earlier sessions focused on U.S. law, we know that AI's use and its impact are not bound by any national borders, and governments around the world are confronting similar legal and policy questions. We've already been participating in numerous conversations with our counterparts and with colleagues from other countries to compare experiences and thoughts. Looking at the global copyright landscape, several questions have begun to emerge. First, how do the international copyright treaties apply to determining authorship and scope of subject matter protection and exceptions and limitations? Second, what actions are other countries or regions starting to take on AI and copyright issues? In what respects are these approaches similar to or different from ours in the United States? Can consensus approaches be found? And, if so, through what mechanisms? And, finally, to the extent there is divergence, what are the international implications? To explore some of these questions, we are delighted to have eight experts with us today.

The first panel will explore national and regional developments in copyright and AI around the world, and will feature four short presentations followed by a discussion among the presenters. The second panel will consist of a roundtable moderated by Copyright Office attorneys which will address both the ingestion and training issues and also the issues of authorship and copyrightability.

So, let me thank our speakers in advance for contributing their time and expertise. And thank you also to everyone joining us online. We had an impressive 2,700 people register for today's webinar. There's clearly tremendous interest in the topic of AI and copyright. And, as many of you are aware, there will be other opportunities to engage with the Copyright Office going forward. In particular, we're planning to issue a Notice of Inquiry on AI and copyright in late August, so very soon, seeking public comments on the full range of issues. I am very much looking forward to today's discussion of the critical international context. So, without further delay, I will turn the mic back over to Chris.

Chris Weston: Thank you, Shira. I have a few housekeeping points before we get started. We are recording today's webinar. The recording will be available on our website in approximately three weeks. For panelists in our second session who are viewing this session, please keep your camera turned off and your mic on mute. The transcription function is activated. The Q&A feature has been disabled as we are not accepting audience questions today. If you are in the audience and wish to share a question or comment with the Copyright Office, we will be soliciting written comments through a Notice of Inquiry later this summer as Shira mentioned. And, as Shira indicated, today's session on international issues and artificial intelligence will consist of two panels with a brief break in between.

Our first panel will feature four copyright experts who will give short presentations on key developments happening around the world followed by a brief discussion. After our first panel ends, we will have a very short break. Our second panel will feature a roundtable discussion with four copyright scholars who will first explore ingestion and training issues, including exceptions and limitations, and other infringement related matters. And then turn to a discussion of authorship and copyrightability. With that, let me turn it over to Maria Strong to introduce our first panel.

Maria Strong: Thanks, Chris. And good day, everybody. I'm Maria Strong, Associate Register of Copyrights and Director of Policy and International Affairs. I'm joined today for this panel by Emily Lanza, Senior Counsel for Policy and International Affairs. Let me introduce our four experts today for this first panel in the order in which they will appear. Peter Yu of Texas A&M University School of Law will provide a presentation on regional developments in Asia, specifically covering China, Japan, Korea and Singapore. Marcus von Welser, a partner at Vossius in Munich, will discuss the European Union's proposed Artificial Intelligence Act, along with text and data mining exceptions in the EU Digital Single Market Directive. Luca Schirru of the KU Leuven Centre for IT and IP Law in Brussels will give an overview of global text and data mining developments in the global south. Shlomit Yanisky-Ravid of the Ono Academic College in Israel will then discuss issues involving the use of copyrighted materials, Israel's approach to AI development and issues related to language bias and discrimination and AI tools and possible solutions. Links to their biographies will be posted in the chat. So, let's get started. Peter Yu will start off. The floor is yours.

Peter Yu: Thank you, Maria. So, there are two different types of issues that are quite important with respect to copyright and AI in Asia. The first one is about copyrightability of AI-generative works. So, within Asia, I think a lot of people would start off with the case in China about Tencent. And this case is about a report on stock market that has been generated automatically by the Dreamwriter from Tencent. What's interesting about the case is that the court found that the report is actually protectable, but not because AI can be an author, but because of the arrangement and selection by the creative team from Tencent of the data input, trigger condition setting, template and choice of corpus style. There is also reports with respect to how AI has been registered as a co-author of an artwork during the pandemic in India. But that decision has been withdrawn a few months later.

In addition to some of the developments I mentioned, there are proposals about sui generis protection for AI-generative work. And, so, for a good example is a slide I show you right here in Korea, where a proposal is actually going through the legislature. A lot of academic industry representatives have also been advocating for sui generis protection of AI created work within Asia. The second issue that's of interest to those of us who think about AI and copyright is about the use of copyrighted works as training data. I think a very good starting point is known as the text and data mining exceptions in different parts of Asia. So, the earliest exception is actually in Japan is developed earlier than the UK and the EU. And, in 2018, there has been a new provision with respect to text and data mining. So, included here is Article 30-4 that talks about the exception that will cover data analysis, computer data processing. And the important takeaway here is that as long as a user is using the work not to personally enjoy or cause another person to enjoy the work that will be covered with respect to text and data mining.

There are also other provisions, for example, about incidental use for either computer exploitation or computerized data processing. Another example that people usually look at is Singapore, which has amended the Copyright Statute in 2021. So, included here is a special exception for computerized data analysis -- computational data analysis. So, there are different conditions. For example, the copy that you use has to be local. You cannot share the copy with others except for verifying results. And there are also definitions -- there are also definitions within the statute that's included, but I think what's interesting is that the illustration actually uses the use of images to train a computer program to recognize images. So, I think that is quite clear, it's about the use in the artificial intelligence context. It's also important to remember that there are also provisions with respect to fair use in Singapore. And, so, the provisions in -- the factors included in the provision is actually very similar to what we get in Section 107.

In China, there's no exceptions for text and data mining. But, during the pandemic, China amended the Copyright Statute. And this is the first amendment since China joined -- the first major overhaul of the Chinese Copyright Law since China joined the WTO. And, so, it's quite important that way. And Article 24, what used to be Article 22, lists the different limitation exceptions. And added to the new statute is clause 13, which says that other circumstances as provided by laws administrative regulations. And it's important to keep in mind that, here, this is not an open-ended clause. It says that the circumstances have to be provided by laws and administrative regulations. So, it depends on what type of laws and regulations have been included. And, at the moment, China still has not released the implementing regulation for the new Copyright Statute.

So, that's something that we can look forward to see whether there will be any special provision with respect to text and data mining. About two weeks ago, there's new Provisional Measures for the Management of Generative AI Services. And I think that's quite interesting in terms of the 21 provisions there. But the important takeaway here is that they only mention that the AI services and also training data has to respect intellectual property rights. So, it's unclear as to how that would affect some of the regulations that are going to be drafted with respect to text and data mining.

And, finally, if you actually go to Korea, it's also important to remember that Korea also has the first provision in -- somewhat different from what we get in Section 107 is that the two steps of those three-step tests are actually built into Article 35-3 of the provision of the -- in Copyright Statute. So, the -- so, these somewhat different from what we got in Section 107. And there has also been reports that new standards and guidelines on copyrights of AI-generated content will be coming out later this year. The four jurisdictions I mentioned here, I think it's important to remember that laws on books can be different from law in action when a law is being applied. But three of the four -- two restrictions I mentioned here are civil law countries. So, law on the books are quite important. The other thing that is also worth remembering is that use as training data, just because the use is actually legal as trained data does not mean that the provision also allows for the use to generate competing works for commercial purposes. So, we will just have to follow the developments in Asia to know how the court is going to apply the law.

So, in the interest of time, what I'm going to end with three closing observations. The first one is that when we think about copyright and AI, we always think about what's the meaning of "author" within the constitution whether AI meets incentives. But I want to encourage people to actually think more about whether there are other policy considerations, because a lot of those issues will be different from country to country. And Asia could be quite different. And one question I love to ask people is, "What if other countries start offering copyright protection to AI-created works? Are we going to change our policy in terms of protection for AI-created works? We have precedents going in both directions. We have harmonizations through a TRIPS Agreement. But we also have resisted the protection of databases. The second observation I want to leave with you is that we are still a little bit early in terms of looking at the future. So, it's possible that we may have convergence in the future -- but it's also entirely possible that we may have divergence in the future. And, at this point, it's very hard to tell. We can try to go -- negotiate international treaties. But, at the same time, it takes time to negotiate international treaties. And, on the slide here, I share the different treaties that have been negotiated since the adoption of the TRIPS Agreement. And the final observation is that we love to separate AI-generated work from AI-assisted work. But, at the same time, in the future, I think a lot of people will be using generative AI to different extent. So, some work will have more human input, others will have more alternative AI usage. And it's going to be quite difficult to actually have a clear-cut separation between the two. And I think it's very important to actually understand more about the interface between the law and the machine, as well as the human creative input. And, so, I'll stop here. Let me pass it back to Maria.

Maria Strong: Thank you so much, Peter, for that very quick regional tour de force. I would like to invite Marcus to the stage. So, we will be going from Asia towards Europe. Marcus, you have the floor.

Marcus von Welser: Okay. So, good morning. My name is Marcus von Welser. I'm attorney at law in Germany. And, today, I will speak about the AI Act and data mining. Let's have a short look on the history of the AI Act. In April '21, the EU Commission published a proposal to regulate artificial intelligence in the European Union. In December of last year, the EU Council adopted its common position. And then, during this year, EU Parliament adopted its negotiation position on the AI Act. In my presentation, I will focus on this -- of the -- on the EU Parliament negotiation position because the two previous proposals did not address foundation models and generative AI.

So, we have a definition in Parliament's negotiation position saying AI systems are machine-based system that are designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations or decisions that influence physical or virtual environments. And the EU AI Act will be applicable if the AI system is placed on the market or put into services in the EU or if the output is intended to be used in the EU, so regardless of where the provider sits -- has its seat.

So, the basic concept of the AI Act is a risk-based approach. We have different risk levels. And there are some applications which are prohibited altogether. For example, subliminal influence, social scoring, real-time remote biometric identification systems in publicly accessible spaces, and so on. We have high-risk systems. For example, certain safety components, certain AI systems in areas such as critical infrastructure, law enforcement or border control management. And then we have the low-risk AI systems and foundation models as a general rule belong to this low-risk AI systems. And we have definitions for foundation models and generative AI already in the negotiation position of the Parliament. And the basic -- or one of the main provisions dealing with foundation models is Article 28b.

And, with regard to generative AI, it says that providers shall [inaudible] document and make publicly available a sufficiently detailed summary of the use of training data protected under Copyright Law. And we have some transparency obligations. For example, natural persons exposed to an AI system should be informed that they are interacting with an AI system. We have a provision regarding deep fakes, [inaudible] shall disclose that the content has been artificially generated or manipulated. And there's an exception for movies and video games with regard to this transparency obligation.

And, with regard to the compliance, there is very valuable research from Stanford looking at whether the big models are already in compliance with the draft. And, unfortunately, at the moment, at least the outcome is it's the -- most of the big models are not in compliance at least with regard to copyright. So, as you can see here, most of the big models, for example, OpenAI or Google, are not yet in compliance with the transparency regarding the training data. Then, finally, we have a recital saying that the new draft, the AI Act, should not affect the already existing provisions in the European Directives. And one of the important provisions is in the DSM Directive.

And I will now address the DSM Directive and the text and data mining provisions over there. So, in the DSM Directive, we have text and data mining exceptions for scientific research on the one hand, and commercial use on the other hand. And text and data mining is defined as any automatic analytical technique aimed at analyzing data in digital form in order to generate information which includes, but is not limited to, patterns, trends and correlations. And regarding commercial use, we have an opt out mechanism.

So, let's have a look. The basic question, of course, is, "Do the TDM exceptions allow the use of training data for generative AI?" And when we look at the history and we find the answer that is not at all mentioned. In the DSM Directive, artificial intelligence is not mentioned and neither, of course, is generative AI. And the focus on that time when the DSM Directive was published was more on medical research, so, for example, detecting cancer, skin cancer and so on. Then unless the -- we have the concept of storage and retention. So, at least with regard to commercial use, it is provided that the training data can be retained as long as it is necessary for the purpose of text and data mining.

And, in the German Copyright Act, we have a provision that training data has to be deleted when it is no longer necessary. And this concept of storage and retention does not really reflect, in my view, the fact that generative AI systems memorize their training data. So, we have various studies from regarding diffusion models on the one hand and large language models on the other hand. And, for example, this study is look at whether the training data can be found in the output. And, as we can see some of the pictures which were included in the training data can be found more or less identical in the output. So, that was the diffusion models.

With regard to large language models, I found it interesting actually from Berkeley researchers who found that in GPT-4 a lot of training data can be found in copyrighted books from -- which are still on the market, like "Harry Potter," for example, and which are copyright protected, of course, can be memorized words for words. And apart from this, there are some legal issues as well with regard to the Berne Convention. We have this so-called three-step test. So, exceptions are only allowed in certain special cases which do not conflict with a normal exploitation of the work and which do not unreasonably prejudice the legitimate interest of the author. And it seems to be that it is controversial whether the interests of the authors are somehow prejudiced by the use of training data. So, there is some litigation cases already going on in the United States.

Then we have the forbidden formalities except -- rule in the Berne Convention, which might be applicable also for the opt out mechanism, because the rights to train only comes into existence when you opt out. Without opting out, there is no real right to train because everybody is allowed to include the works. And, finally, this is more a European issue. We have the problem that, in the DSM Directive, there is no compensation for data -- text data mining. And it is questionable whether this is in accordance with the fundamental rights of the European Union. So, in summary, the AI Act takes a risk-based approach. It does not affect existing copyright directives. And the only provision regarding copyright is that providers must publish summaries of copyrighted data. And it is controversial until now whether the TDM exceptions apply to generative AI. Thank you very much.

Maria Strong: Thank you very much, Marcus, for your excellent summary of the very complex goings-on in Europe. We are going to take a little tour now towards the south and I'm going to invite Luca to the stage to continue the discussion on the work he's done in researching text and data mining exceptions in the south. So, Luca, you have the floor.

Luca Schirru: Good morning, everyone. My name is Luca and I'm Researcher and Executive Director of the Brazilian Copyright Institute and a postdoc researcher at CiTiP and KU Leuven. First of all, I want to congratulate the U.S. Copyright Office for the organization of these important events. And thank you for the opportunity to be here providing a brief overview of how countries in the global south regulate text and data mining in the copyright laws. In order to make the time, our presentation we focus on Latin America. However, it is worth noting that there have been important developments in WIPO committees both in the SCCR and the CDIP with the presentation of a pilot project on TDM to support research and innovation in Africa, and the proposal for a draft work program on exceptions and limitations, both presented by the African Group.

So, a recent study mapping the research limitations and exceptions of the national copyright around the globe illustrate the difference in how research rights are treated in the global south and the global north. In the proposed classification, while green countries provide a more permissive approach to research by allowing multiple uses of all kinds of works, red stands for the opposite side of the spectrum, usually picturing countries that only allow the use of excerpts for research purposes. Which, for TDM, can be prohibited. In this method, it's clear that while there is a significant concentration of green countries amongst the ones commonly referred as part of the global north, the red ones are mostly present in the global south, as is the case of Africa and Latin America.

A different working paper that tracked change in the national laws over the last 21 years confirms the distribution seen before. Created from publicly available data in 2021, it represents an accurate snapshot of recent international copyright history. Instead of working with colors, this article's grading scheme was based on numbers according to how restrictive the law is. Zero stands for the most restrictive, the red countries. And five for the most permissive laws, those countries highlighted in green. And, as we can see, high-income countries usually enjoy more permissive laws, which has been consistent for the past 21 years. On the other hand, by going to the details on the countries that increase and decrease their attributed scores, we see that amongst those who increased it, there are only two African countries. At the same time, they are the majority of most countries with decreasing score with a total amount of six occurrences. Concerning Latin America, and using this very interesting interactive map that allow us, for example, to filter for text and data mining, limitation and exceptions, we noticed that the region is even more restrictive than we considered before when addressing broad research exceptions.

The current status of Latin America is that almost all countries don't even have a TDM exception. And, even for those who have research exceptions, these may be limited to the use of excerpts of small portions, which may be way less than optimal when using TDM techniques. Starting 2020, initiatives amongst -- aiming to promote legislative amendments in Latin America may eventually change the scenario. In July 2020, changes were proposed in Mexico by adding a limitation for reproductions and extractions for the purpose of text and data mining when the user has had lawful access. Also in 2020, Uruguay has proposed to allow for the reproduction of works for computational analysis, but only if carried out within the scope of non-commercial research. Since 2022, and even though it does not specifically address TDM, there is a view under discussion in the Peruvian Congress that may change the National Copyright Law to authorize the reproduction of work stored in the cloud for private users if there is a lawful access to it. And, more recently, in 2023, after a series of public discussions involving different stakeholders, an AI bill is being discussed in the Brazilian Senate.

Briefly stating the article of the bill that addressed text and data mining, authorized automated use of works, such as extraction, reproduction, storage and transformation in text and data mining process, but is limited to the activities that are being carried out inside research and journalism organization, as well as museums, archives and libraries. The debates carried out in the scope of these experts commission proposed by the Brazilian Senate and also collected from the contribution of different stakeholders provide some important insights. The first one is related to the need for further information on the technical aspects of TDM since it may not even be inside within the scope of copyrights. Also, some stakeholders show the more flexible and innovation-friendly approach to issues like liability, for example, while keeping a more conservative or a more proprietary approach when it comes to copyrights. And we also noticed that traditional copyright industries, for example, the audiovisual and the publishing industry, they are more resistant to TDM limitations than those companies that are involved with technology.

Finally, it's important to stress that this presentation was limited to a brief overview on the regulatory framework of these selected countries. Therefore, issues like using TDM from known research purposes or other techniques employing for training AI systems were outside the scope. Generative AI brings an additional layer of complexity and requires to draw important lines between concepts and scenarios that may be easily blurred. The first one is the differentiation between text and data mining and the training of generative AI system. While recent studies reveal that TDM may be present in the overall training of AI systems including generative AI, from a copyright perspective, it does not seem reasonable to treat each and every use of works for AI purposes equally. And this would be because as usually seen in research, for example, employing TDM techniques to extract patterns and correlations from a large amount of data does not seem to be a copyright issue as facts and data by themselves are not even protected. At the same time, the training process and the desired outputs obtained from the use of generative AI systems may be different as the system can provide outputs that may potentially compete with the original works used in the training process. And understanding which kind of users are needed within the training process and whether these users are deemed exclusive, right, inside, for example, the scope of the reproduction rights, is an important first step. Another important consideration, and this will be the last one, is that copyright today is seen as much more than the set of rules governing the use of original expressions.

By defining the scope of what can be used to train AI and in which circumstance, copyright may restrict and potentially hinder the development of the country's AI industry and, more broadly, its technological development. Therefore, considering the clear public interest related to the regulation of TDM, which extrapolates the interests of parties involved in a private deal, the needed use for the purpose of research should be shaped in a way that cannot be overridable. In addition, regulation must also consider these biggest players on the market. But, mostly, the new entrants and those players who are central in the flourishing of national systems of innovation. While the biggest players have a relevant impact in the overall public, they have already benefited from a much less regulated framework to train their system. And a restrictive regulation may mostly affect the development of new business and perpetuates the concentration effects. Finally, we now have a unique opportunity to redesign the copyright system as a tool for innovation in a much different and technological context where exclusiveness and property are not the only, nor the better, solution to regulate that intensive technologies, especially when we consider different cultural, technological and economic aspects of different countries and jurisdictions. Thank you for your time and attention. And I'm looking forward to learn with the other presentations and debates that will follow. Thank you so much.

Maria Strong: Thank you, Luca, for sharing the results and the ongoing news of the survey work that you and your colleagues are doing, as well as the remarks you shared on some of your views on the way TDM is going along. And, to continue our conversation here today, I would like to invite to the mic Shlomit to talk about copyright language biases and language barriers. The floor is yours.

Shlomit Yanisky-Ravid: So, yeah, good morning, everyone, good evening and good afternoon for the [inaudible] different audience from different parts of the world. First, many thanks to the Copyright Office and especially to Shira and Chris and Maria and Emily and Danielle and Alicia for organizing these wonderful open arms to the international overview event, and of course for having me. I've been -- I was researching AI and IP for many years. And now I'm happy with a chat on these advances that everyone see that -- yeah, and can understand what I was speaking about. So, I am a professor of law at Ono Israel. But, besides being at Ono in the winter -- in the -- yeah, part of the year, I teach at Fordham from 2012. And I am affiliated with the Yale Law School from 2011. And I just want to say that all the -- this presentation was all made by artificial intelligence. The content is mine, but all the other was made by AI autonomously.

So, I would like to speak about several thing, and the importance of AI tools, language barrier, solution to copyright issue, the Israeli approach and some more. So, I don't think we should elaborate more on the importance of -- yeah, for promoting science and useful arts and creative industry in all fields, and how AI adoption is a key driver for advancement in creative industries, such as software industry, entertainment, design, music, science and many more industries. But, beyond the signs and patterns, like using it for manufacturing and developing drugs and other creations. So, creative content generate and automation is really highly used in creation and creative industry. So, I mean -- but the -- yeah, the picture is not so bright.

And I want to focus first on the barriers -- the language barriers and the disadvantages for non-English speaking countries. So, AI systems are mainly trained in English and, therefore, non-English speaking countries suffer from disadvantages and inferiority and limitations. The access to advanced -- yeah, advanced science, art, tech, studying, access to knowledge and, generally speaking, being part of your progress and being part of the society. And we have several international treaties that mentioned that. So, I mean, if someone doesn't know English and he cannot really use the chat and other creative AI tools, so that means these people are with disadvantages. You can see here the English-speaking countries in green. And, here, it's -- the green is English-speaking countries in the map below. And the red is non-English speaking countries. But because like they're English-speaking population -- so we have like the red -- the green is English-speaking population, the majority of the country. And the red is just a few percentage. But most of the countries in the world today, they don't speak English. And users and AI software developers and creative industry have become inferior. Just for example, Voice Assistant cannot understand. And I'll speak about translation soon.

The impact on global development and disparities between developed countries and developing countries is critical. And the gap increase because the English-speaking countries are the first takers for AI industry and from promoting science and technology and creative arts. So -- yeah. So, that's I think something that we should think about globally. So, one solution is that -- it's AI translation tools, such as Google Translate, but we have several more. So, I think we all remember that or we all know or we heard about the Tower of Babel. And that's how I pulled one of my articles from the Tower of Babel to Google Transfer -- to Google Translate. So, in the power of the -- in the Tower of Babel, like your kind wanted to speak in one language. So, they build like a total, yeah, tower and God punch them according to the Bible. Was it a punishment or was it just to promote diversity? So, Google Translate and other translation are, yeah, vastly and rapidly used by everyone.

So, like Google Translate itself processing a billion of transactions. So, there are a lot of benefits that we all know that we can communicate with others. But, if you come from a non-English speaking country and you want to use these tools, so there are a lot of drawbacks and challenges, cultural influences, sensitivity and nonsense and many more. I'll share this slide so you can follow that. And, also, as -- yeah, as we know in copyright that using translation means that it's a derivative works, that means that we infringe the right of the original author. And, therefore, I mean, we are actually everyone using Google Translate. But Google Translate is just an example for all AI tools when we use copyrightable works, when we use -- without an authorization. So, we are violating the author rights. And, therefore, we are subject -- we might be subject to legal sanction if it's not fair use. I mean, we might -- should -- we should ask for a license, pay for the author and, you know -- yeah, and the author should get benefits as well. And the moral right and the credit is also an issue. So, we have like a tension between authors, the public and use of right -- and the benefits of using the artificial intelligence tools for -- by everyone.

The other thing is AI bias. So, bias in AI language models can manifest various forms such as gender, racial, cultural, insensitivity leading to inaccurate and unfair output and perpetuates stereotypes and exclude group of people. Just for example, I asked the AI, I put the prompt to create some -- you know, some of these figures. And you see very stereotypes figures are being designed here. And that perpetuates -- yeah. And biases in translation are also because some countries, such as Israel, have like gender differences between when we speak like French as well. So, translation is really inaccurate. So, yeah, I will try to go really, really quickly. So, AI in copyright material, so it's not just the fact that we, all the public, are using AI for translation and for any other, like the chat, the Bard and all these other tools. And we put copyrighted works. And we use that without permission.

But, also, the data mining is an issue. So, there is a conflict between the right of copyright authors and other stakeholders and justification for copyright. That's, at least, the economic perspective, incentive the human creation, personal, labor Rawlsian and also the distributive justice. But we also have, yeah, very many interests of the AI industry to promote science to advance. Here, you can see an example of a website that said that they can -- you can create like songs of famous people. And, I mean, there is [inaudible] to voice itself can be considered as copyrightable. Solution for data mining and copyright issues may be the possible solution I think it's like to [inaudible] between copyright holders and -- copyright holders and the industry. So, we can think about obtaining license or obtaining blanket license from artists' organization and collaboration, establishing partnerships. So, I think there should be solutions. I'll just really just mention the Israeli perspective.

So, the Israeli State of the Minister of Justice and legislation and delegate counsel for civil law, they published opinions recently. And they -- their opinion actually said that the machine-learning training and AI is generally permitted under existing copyright doctrine of fair use basically, which is based on the U.S. doctrine. However, the one thing that was forbidden was work of the same artist -- of a single artist to compete with their market. So, just to speak generally, right in works of generated -- right of work generated by AI. So, I know there is a gap between different perspective. And I just hold this perspective that it's the wrong question of should we authorize --

Emily Lanza: Hi, Shlomit, sorry to interrupt. Can you just wrap up, please, thanks, so we have enough time?

Shlomit Yanisky-Ravid: All right, all right. So, I'll just conclude and say that, yeah, I mean, there must be a solution and we propose some solutions. And I'll -- yeah, I'll publish the slides. And sorry for taking extra time. I prepared so many things to say. And, yeah, thank you.

Maria Strong: Yes, thank you, Shlomit. There are a lot of things to say indeed. And I'm just going to say thank you to the panel. And I want to turn -- we have a little bit of time for questions that Emily might be able to prompt. Thank you.

Emily Lanza: Thanks all for your excellent presentations. Just to kind of wrap up here, I'd like to conclude with one question. So, do you expect the various national approaches to regulating copyright and AI to converge? If so, will or should this happen through a process of interpretation and formal harmonization? Or will a treaty be called for to address differences in national approaches? So, I guess the challenge here is to answer this weighty question in just a few seconds. So, who would like to take the first stab?

Peter Yu: So, let me start. So, I think trying to separate convergence and divergence is probably not the best approach. I think we'll converge in some areas, but we also diverge in some areas. I think the big question is how are we going to set international standards. Would that be through treaty? Which will take a lot of time. Or would we want to go for software recommendation? And the final thing is that based on Luca's presentation, it seems like the big players are the U.S., EU and Asia, right, in terms of a lot of the AI TDM issues. So, if we're going to open the discussion, then it will become more complicated for other countries that have not yet reached that stage.

Emily Lanza: Thank you, Peter. Anyone else?

Shlomit Yanisky-Ravid: Yeah, I think like an international treaty may be the solution. I think IP roots are with international treaties and like, you know, the Berne Convention and others and the TRIPS. So, therefore, I think because even though there are differences between the countries, I think there should be some kind of like agreement that everyone should follow because AI is above nations, has like international characters.

Emily Lanza: Thank you. I think we have a few seconds left. So, who would like the last word for this panel? Anyone?

Luca Schirru: Well, there is no -- not much to add after these very nice interventions. But definitely one thing that we need is some kind of harmonization because these -- when we are working with data, sharing data, training AI systems with data, we are talking about research projects or businesses, they are definitely go over the national borders. So, the ins -- in any kind of manner, which one exactly we still don't have an exact answer maybe. But we should address this cross-border uses, especially in the digital environment, using -- for research and education.

Emily Lanza: Marcus, I just want to give you a chance if you have any thoughts on this question before we wrap up.

Marcus von Welser: Regarding what?

Emily Lanza: The question I asked whether you think the informal or more formal approach when these issues converge.

Marcus von Welser: So, of course, I would -- an international treaty would be desirable, but I think the political and economic interests of the different countries are different. So, for example, the United States, you have a different approach let's say from China. So, it might be difficult to reach a common understanding. But with regard to the copyright provision, we already have an international treaty, such as Berne Convention. And there's also these mechanisms in the WTO, the TRIPS Agreement. So, maybe this might be a solution. For example, if you go to the three-step test, which in my opinion is very important in this scenario.

Emily Lanza: Great. Well, thank you all for answering that question so succinctly. I know it's a big topic. And I'm sure we'll have opportunities to discuss that question and answer it in occasions to come. So, I'll turn it back over to Maria.

Maria Strong: I think with that I'd like to wrap up and thank this panel. We're going to take a short break. And I think Chris is going to lead us into that.

Chris Weston: Thanks, Maria. Again, we'd like to thank all of our experts for participating today. For those of you in the audience who will be staying on for the next session, we invite you to take a five-minute break. We will resume at 11:57 Eastern time with our second panel. It will feature a round table discussion with copyright scholars who will first explore ingestion and training issues, including exceptions and limitations, and other infringement-related matters. And then turn to a discussion of authorship and copyrightability.

We are joined by a very distinguished group. They are Jane Ginsburg of Columbia Law School, Andres Guadamuz from the University of Sussex, Bernt Hugenholtz from the University of Amsterdam and Matthew Sag of Emory University School of Law. Full links to their bios will be posted in the chat. However, before we start, I inadvertently omitted some housekeeping points. So, once again, we are recording today's webinar. The recording will be available on our website in approximately three weeks. For panelists from our first session who are staying on for this session, please keep your camera turned off and your mic on mute. The Transcription function is activated. The Q&A feature has been disabled as we are accepting audience questions today -- or as we are not accepting audience questions today. If you are in the audience and wish to share a question or comment with the Copyright Office, we will be soliciting written comments through a Notice of Inquiry later this summer. So, once again, I would like to invite our distinguished panel of experts. I should mention that we are having some connection troubles with Jane Ginsburg. So, hopefully, she will be able to join us as soon as she is able. So, to get started, we are interested in learning more about ingestion and training issues, which includes exceptions and limitations and questions about infringement. To begin, I will turn to Danielle who has the first question. When you are ready to speak, please use the Raise Hand function and we will call on you. So, Danielle, the floor is yours.

Danielle Johnson: Thank you, Chris. As we heard earlier, copyright treaties include a three-step test which limits exceptions to copyright to certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author. In your view, how does the three-step test affect exceptions that might apply to the development and use of generative artificial intelligence? You can raise your hand or use the Raise Hand feature if you'd like to jump in.

Matthew Sag: Thank you, Danielle, for the question and thanks to everyone of the Copyright Office for all the great work that you're doing here. My perspective on the three-step test is that I think that in the United States, as long as the Fair Use Doctrine is applied correctly, the three-step test is not a significant problem. I think the text data mining that is a non-expressive use and training for machine learning that also fits within that sort of definition of a non-expressive use, they are outside the normal exploitation of the work. They are uses that don't involve transmitting original expression to a new audience. And even though they're at a large scale and they're spectacular, they are well outside the traditional parameters of Copyright Law. So, I don't see the three-step test as being a major impediment, at least to developments in the United States.

Danielle Johnson: And Bernt?

Bernt Hugenholtz: From an EU perspective, which is quite different because we have specific exceptions in the books, in the Digital Single Market Directive we have two provisions on text and data mining which clearly apply in the AI ingestion context. I still agree with Matt's analysis. I do not see the three-step test play a very large role in interpreting these exceptions because they're very specific and they're also very recent. Although the three-step test does apply, the DSM Directive, in fact, reminds us that it does apply even in the context of these exceptions since they are so specific. There is -- if the facts fit the exception, we have a certain special case. There is a mechanism to avoid a new exploitation which is not normal since there is an opt-out mechanism in place. So, we actually have the various bases of the three-step test already covered in the European context. That doesn't mean we have no problems interpreting these exceptions. But I don't see a role specifically for the three-step test.

Danielle Johnson: And Andres?

Andres Guadamuz: Yes, I agree as well. I think that the three-step test may come to play at some point at a national level. But, at the moment, I don't see that it's going to apply at least with the exceptions that we have. Potentially with the broadest exceptions, like the Japanese one which has been in the books for quite a while. So, it's always good to remember that their exception has -- even if it's quite broad, it's been around as Peter was mentioning in his presentation. Eventually, I think that the broad exceptions may be tested in court. And I don't think that there is any stomach right now in the international community to test the application of -- or the exceptions at an international level. The TRIPS system in that respect hasn't worked for over 20 years. So, I think what we could see eventually with some exceptions that are considered to be too broad, I agree with Bernt completely that I think that the European exceptions are within the parameters of the three-step test. And, also, with Matt, the U.S. as well. I think that a broad exception potentially could get challenged.

So, for example, the UK has an exception since 2014 which is for scientific research. It hasn't been tested in court. And I think it's well within the parameters. Now, we had a very broad exception proposed last year by the UK IP Office that was bent eventually because there was a little bit of political aggro from the creative industries on that one. But, interestingly, I was thinking if this went through, it may have been challenged at some point in the courts. There is precedent for this. The -- back in 2015, the British Academy of Songwriters, Composers and Authors took the -- asked for a judicial review on an exception for private copying that had been created in 2014. And they won. The exception was removed. That means that whenever we're doing any private copying in the UK, we're infringing copyright. But, yeah, it was brought down and it was brought down precisely -- big part of the consideration at the time was that it was in violation of the three-step test, that it didn't comply with -- it didn't compensate the musicians and composers. So, if we ever see any broad exceptions presented again, potentially this could be challenged in the courts. But, at the moment, I think that the exceptions as they are written are going to remain. Whether that is going to be the case in the future, we'll see.

Danielle Johnson: Thank you all for your input. I'll turn it over to Chris for the next question.

Chris Weston: Thanks, Danielle. So, following on that for the topic of exceptions and limitations, some countries, such as Japan and Singapore, have broad exceptions in their laws that may be interpreted to apply to training AI models. The EU's 2019 TDM exceptions try to strike more of a balance among the different types of users and uses. And other countries, like the United States, have open-ended fair use provisions. What are the key challenges you see for each of these approaches?

Matthew Sag: I think that it's important to see these different approaches as embedded within different legal cultures. So, I don't think we should be too worried that in the United States we hand a little bit more over to the courts, and in some civil law countries they try and be more prescriptive. And, obviously, there are, you know, swings and roundabouts there. The EU approach of basically having different rules for non-commercial researchers working at universities and cultural institutions, we haven't really seen that cashing out in U.S. case law. Although it's possible that it could. It's possible that with respect to some aspects of generative AI the fact that it's a commercial user might have some impact when you're

looking at the fourth factor. But I think the difficulty in Europe and in the U.S. for making a hard distinction between commercial and non-commercial is what do you do about non-commercial research funded by commercial entities that's then used to train general foundation models that are then fine-tuned by commercial actors. You know, you might think, you know, that's a great thing, and I would mostly tend to agree. But it can also be seen as a sort of work around between this attempt to draw a line between commercial use of the non-commercial uses. So, I think that's an interesting challenge that we need to think about a lot more.

Chris Weston: Bernt has raised his hand as well.

Bernt Hugenholtz: I think a clear challenge to the EU system which has this double exception, one for non-commercial research and one for all other types of uses, commercial or non-commercial, journalism, AI, what have you, is it is particularly in the second one, the Article 4 of the DSM Directive, TDM exception, which has this unique feature of an opt out, a way of balancing supposedly the interests of the right holders with those of AI developers. If you don't want to be trained, you can opt out. And on paper, that's a fair balance indeed. The devil is in the details. What does an opt out mean? This is a simple question which is -- and will remain unanswered for quite a while. Does an opt out have as its consequence that the specific digital object which is out there on the internet cannot be mined, or does it have a much more general effect on the work that is represented in that digital object? Can you, in other words, opt out your oeuvre or your work generally from use by text and data mining operations, by AI training? Or do you have to do it one by one by one for all the objects, all the duplicates of your work that are available out there? This is one of many questions that has been -- have been raised by the opt out mechanism. And they are very difficult to answer. It will take quite a while before we have clarity there.

So, I see the main challenge in the EU approach in basically interpreting how this opt out is going to work. There are other issues as well. And, as you know, it can take quite long before the European court has made such decisions as the final authority on EU law. We will have national decisions earlier than that. But it could -- even that could take a while. There are no cases before the courts in Europe as far as I know. In the meantime, obviously, we don't really know what this opt out really has in terms of legal effect. Another problem is we don't know what a standard opt out looks like. The directive wants -- or tries to promote a model where opt outs are machine readable, automatically recognizable by scrapers. As I -- as we speak today, there is no industry consensus on what that means. So, there are many practical issues with the way the TDM exception is to operate in the years to come.

Chris Weston: Thanks very much, Bernt. Before I go to Andres to answer this question, I want to welcome Jane Ginsburg, who has managed to join us via phone. And, Jane, we will loop you into the discussion as soon as Andres is done. Andres, please go ahead.

Andre Guadamuz: Yes, just quickly, Bernt, there is a case in Germany right now. There is a photographer they're suing LAION the makers of the -- one of the datasets. So, it may end up in the Court of Justice. We don't know. But that is the only one I know in Europe. Regarding the challenges, I think that the challenges when it comes to this, as what happens with all exceptions and limitations, is to try to strike a balance. And I think that the balance here is quite tricky because some countries are seeing the development of artificial intelligence as important for their economy of the creation of in -- creation of industries that are taking advantage of the artificial intelligence revolution, let's call it like that. And apologies for using that word. And I think that each country is going to have to revisit. Some exceptions were created, and I think that's something that came up in the first panel, that some of these exceptions were created with different objectives in mind. They were created to cure cancer to -- for medical text and data mining so that we could have new discoveries. And, so, I think that the challenge is going to be precisely to look at the different interests at stake, the interest -- the economic interest of fostering innovation and potentially bringing these companies that are going to generate revenue for the country. In the case of the UK, there is a big example that Brexit -- I'm going to shock some people, but Brexit is not going that well. So, some countries may see this as a potential issue where they can try to take advantage over their neighbors. On the other hand, there is the legitimate interest of creators to object if there is any information that is being used to scrape or input data. So, again, it's always balanced. I don't know how we're going to square that circle. But, yeah, it's going to happen.

Chris Weston: Thanks, Andres. Before I turn it over to Danielle for the next question, I want to give Jane an opportunity to weigh in on the first question, which she did not have an opportunity to hear. Jane, basically, the question was, "How does the three-step test affect exceptions that might apply to the development and use of generative artificial intelligence?"

Jane Ginsburg: Okay. Well, it only took an hour and 15 minutes to get connected. And I thank my much more device-savvy daughter for finally figuring out how to get me on. As for the question, I think the third step is a very interesting one. The -- I don't know what the others have said about the three steps. The first one, of course, it's not certain special cases if it's all works.

It's -- it may compete with a normal exploitation of the work if licensing works in bulk for text and data mining and AI training is a market. And then the third step I think is quite interesting because, unlike the fair use fourth factor, the impact on the copyrighted work, the third step talks about the legitimate interests of the author. The Berne version of the third step, not the TRIPS version of the third step. So, Berne, WIPO Copyright Treaty's version of the third step. And there's an interesting question whether the legitimate interests of the author are prejudiced if the ingestion of her works into AI training has the result of competing with her ability to earn a livelihood, because it substitutes for the work that she might be doing, not necessarily for any individual work that she has done, that would be a U.S. fair use inquiry, but whether it substitutes for her livelihood as an author by creating works in the style of. Normally, we don't think the style of is protectable. But if there is infringement on the input side because the works are actually copied into the system, then there's a very interesting question about whether the outputs do prejudice the reasonable interests -- the legitimate interests of the author.

Chris Weston: Thank you, Jane. I will turn it over now to Danielle for our third question.

Danielle Johnson: Thank you. And sort of continuing down that infringement path, earlier we had talked about the debate regarding the ingestion of copyrighted materials to train AI models. Is that infringement? Or is there an applicable exception? And if we are assuming for the purposes of discussion that infringement is involved, who in the chain of creation of the AI-produced output would be liable? I see a lot of fast hands now. I'll start with you, Andres.

Andres Guadamuz: Yes. Sorry, wanted to break the monotony of everyone following the same order. It depends. But most -- for the most part, there is going to be some reproduction. At some point, ingestion cannot happen without -- or some form of reproduction. So, if we consider that that, obviously, has taken place, and I'm saying every case is different, but I haven't seen a model that has been trained without some form of reproduction at some point. The production -- the reproduction can be quite quick. There is something -- some preparation. But let's assume that there is almost always going to be reproduction. Now, that is where the exceptions come in. I am of the mind that, for the most part, the reproduction is just for the training so the model can have some mathematical statistical data on the inputs. Sometimes it's billions and billions and billions of works. And, in the case of images, can be five billion. And the -- so, is that -- are there exceptions? I think that we'll need to find out. I think that the most likely exception that falls here is one that we use all the time, the internet exists because of this, the transient exception -- the transient copy exception for purely technological processes. I won't go into a lot of detail. But I think that this is going to be a tricky one. It may depend on how much of an independent economic value does the output have for the creators. I have no idea. But I just think that this is probably one for the courts. And I think that, to me, the -- apart from any text and data mining exceptions, the most likely exception that is going to be argued is transient copying.

Danielle Johnson: Thank you. Bernt?

Bernt Hugenholtz: I agree that transient copying, which is specifically excluded under the EU copyright rules already since 2001, can play an important role in this analysis. Of course, we have the text and data mining exceptions that were already mentioned also in the first panel. I just want to make one other observation here. If we're looking at the chain of possible culprits here, assuming there is infringement, if these exceptions -- assuming these exceptions for some reason do not apply, I think we should have a close look at how these training data actually end up in the model. There is a lot of scraping going on even before any AI training is involved. There are actually special companies doing that and offering off-the-shelf training corpuses. So, if you're looking at potentially liable actors, you could be looking there as well. But, again, I think this is a bit of a theoretical exercise given the fact that, at least under EU standards, there's a very good case to make that the training is actually an exempted act under these various European provisions.

Danielle Johnson: Thanks. And, Jane, your thoughts?

Jane Ginsburg: On transient copying, I don't think that the AI training data would meet the criteria of the Article 5(1) of the EU Infosoc Directive. And I think Bernt said that sort of. For the U.S., we don't know how transient is transient, what is a period of more than transitory duration. We know from the Second Circuit that it needs to be more than two seconds. It -- whether it needs to be how many -- how much more than two seconds or minutes, we really don't know. So, it's not clear under U.S. law whether a transient copying approach would apply. It's not an exception because if the copying is too transient, it doesn't count as copying. As for other applicable exceptions, I don't think, back to the Berne Convention, that the quotation exception would apply to training data because the quotation exception presumes that there will be a disclosure of the quoted material to the public, which is why there is an obligation to disclose the source and the name of the author if possible. None of that would make sense if the copied material were not then subsequently divulged. But if we're just talking about the input stage, I don't think we have something that would count as a quotation under the Berne Convention exceptions. And I'm -- I won't dwell on fair use at this point, because I imagine a lot of people have a lot of things to say about fair use. So, I'll hold off to a more specific discussion of fair use and inputs.

Danielle Johnson: Thank you. Matthew?

Matthew Sag: I think that both in the EU and the U.S. there's no way this falls under transient copying. You know, if you actually look at the mechanics of how you -- like how machine-learning training works, like people aren't storing files or parts of files for anything you would measure in seconds or fractions of seconds. They're storing them for months. Right? If you are going to train a model, you're often going to be thinking, "Well, I'm going to have to retrain this in six months' time." So, it would be ridiculous to -- just to erase everything. If you're concerned about bias or model learning from inappropriate sources, the only way you can do what you need to do is to actually have a semi-permanent copy of the training data and then clean that up. So, you know, if you're training an LLM, you want to take out all the hate speech. Right? And, you know, that is like quite a permanent activity. It's not transitory. So, it doesn't really work.

Also, when you're training a model, deduplicating the training data is incredibly important in order to avoid infringing outputs. And that is like, you know, a difficult and sort of part science/part art process that's going to take a lot of time. So, I just don't think transient copying is -- yeah, unless I misunderstand the law in Europe, but certainly not in the U.S. that there's just no way. Some people have raised implied license. And I guess that is possible under certain scenarios. But I think, you know, in the EU, Article 3 and 4 are basically where it's at. And in the U.S., you know, we actually have a long line of cases that are at least conceptually similar, you know, the non-expressive use cases, reverse engineering, plagiarism detection, HathiTrust, which was very much, you know, machine learning in the background. And that was the point of creating HathiTrust and Google Books. So, you know, that -- you know, that is the landscape as I see it.

Danielle Johnson: Thank you. And I will pass it over to Chris for the next question.

Chris Weston: Thanks, Danielle. Do you expect the various national approaches to regulating copyright and AI to converge? If so, will or should this happen through a process of interpretation and informal harmonization? Or will a treaty be called for to address differences in national approaches? We asked this for the first panel as well.

Andres Guadamuz: A very short answer, no. It hasn't happened with anything else that we've seen. Harmonization has been pretty bad in almost every aspect. I don't expect this to be similar. I think, as I mentioned earlier, some countries are actually going to use this as potentially trying to get one over their trading rival. So, yeah, the answer is "no."

Chris Weston: Okay. Short and sweet. Bernt?

Bernt Hugenholtz: I agree. This is not going to happen anytime soon. And if it's going to happen, it will be way too late. Look at how long WIPO's taking for something mundane as protecting broadcasters, they've been doing that -- trying to do that for the last 25 years to no avail. And, more generally, we haven't been harmonizing exceptions at an international level except for the three-step test, which has found its way in various international instruments. And there are some fragmentary rules on other exceptions. But an international treaty is not going to happen. Then the other -- the alternative could be some international consensus in an informal way. Before that happens, we'll have to have clarity about what the courts think the law is in the various dominant regions. They're fair case -- fair use cases now before the U.S. courts before that has been finally cleared up, will take a couple of years. I can't imagine that before that international negotiations will take place. What we're already seeing now are quite different approaches. As was said in the first panel, there are also clearly diverging interests. There are countries that have AI champions and expect a lot from AI technology. There are other countries that are behind and might be more inclined to protect authors' interests. There are regions that want to do both at the same time, like my own region. It will be very difficult for many reasons to come to a consensus, certainly, in a formal way.

Chris Weston: Okay. Thanks, Bernt. Matt had his hand up.

Matthew Sag: Yeah, I don't hold up much hope for an international convergence driven through any kind of treaty, just because of the pace at which things are developing at the moment. I think that there might be economic forces that drive towards convergence in fact, if not in law. I think that lawmakers in the U.S. are very concerned about driving technology overseas. I think that is probably gonna restrain some countries and drive other countries to adopt more liberal policies than they might otherwise do. And one thing that's going to be driving this is that, you know, this isn't like another market where you can quarantine national markets. Once an LLM is trained or a machine-learning model is trained, you know, those models themselves don't contain the training data, despite common misconceptions to the contrary, except in very exceptional cases. They don't contain the training data, so there's no real barrier at the moment in training in a low-protection jurisdiction and exporting to a more restrictive jurisdiction. And, so, I think that might, you know, just pragmatically drive some convergence. And I think some of you might see that as a race to the bottom. You know, views will obviously differ on that. But I would say, you know, I think we're seeing some interesting litigation in the U.S., people sort of testing, well, what are the limits of fair use. I don't think that we're going to see a U.S. court saying that training for machine learning is never fair use. Right? That would contradict a lot of established precedent and, you know, years of industry

reliance. But we might see more restrictions on commercial actors. We might see some pressure on, you know, avoiding sites of known infringement. And, you know, I think U.S. courts might look to the EU and concepts like lawful access and the sort of -- the division in Article 3 and Article 4, and they may take some guidance. I'm not saying one way or another whether they should, but I think they may. So, we may -- you know, like we may still see some convergence towards a higher protection standard, you know, through the evolution of U.S. case law. But, yeah the future is quite uncertain.

Chris Weston: Thank you very much. I'll turn it over to Danielle for the next question.

Danielle Johnson: Thank you. We will pivot slightly to licensing. And, so, is a licensing regime, whether voluntary or compulsory, a feasible option for remunerating artists whose works are used to train generative AI models to the extent that an exception does not apply? And if you have any thoughts on how the compensation model would work to provide meaningful remuneration, we will take that as well. Thank you.

Matthew Sag: I want to say, no, I don't think that will work, and I don't think that it should work. I don't think that it will work because models are now being trained not just on images and works made by people, but on images and texts made by models that were themselves trained on things made by people. And, so, the sort of -- the tracing forward of, you know, who is entitled to compensation for what, I just think is incredibly difficult. And I also don't think that it should work, because I think that I wouldn't want to subject The Wall Street Journal or The New York Times to any kind of compulsory license. I think that, you know, the -- you know, the sort of copyright-friendly future for generative AI that I see is foundation models trained on things that are genuinely available on the open internet respecting opt outs, respecting, you know, do not train tags, respecting paywalls. And then those models being fine-tuned on high-value resources or, you know, more customized models being created based on special sort of high-quality content. And that -- you know, that we should leave to the free market. I don't think that any sort of general license scheme, especially not a compulsory license, I just don't see that as working or advisable.

Danielle Johnson: Thank you. Jane?

Jane Ginsburg: Yeah, I'm not a fan of compulsory licensing either. Although it may be implicit in Article 4 of the EU Directive because that provides for an exception or a limitation. And limitation generally translates to compulsory license. So, I'm wondering if Bernt has any thoughts on that. As for free market licensing, assuming that there were a way to organize the various authors so that they could effectively collectively license for this purpose, I agree that that would be desirable. Whether that requires antitrust flexibility in the United States or, for that matter, in the EU, I would leave to somebody who knows a lot more about than I do.

Danielle Johnson: Yes, Bernt?

Bernt Hugenholtz: As for Article 4 of the DSM Directive, I don't think that is a compulsory license. It's an exception or an exception as limit -- or -- and limitation as it is written in the books. But I don't see a big difference between the terminology. It is an exception and it doesn't in itself require a payment as what would be normal with a compulsory license. Having said that, there has been some -- there have been some commentators in Europe suggesting that the opt out model, which is built into Article 4 of the DSM Directive which allows rivals to opt out from being scraped, from being trained, could be used as a leverage for a licensing scheme. This is still very theoretical. And whether this is actually feasible is a big question. One problem with licensing solutions would be that you would be looking at not the usual professional creators only and their copyrights, but with copyrights of everyone who has ever contributed anything original in terms of copyright to the web, because all that's being scraped, including tweets and very mundane content that under European standards qualifies for copyright nevertheless. So, if you would base -- create a licensing scheme, you would have to look at means of remunerating millions, tens of millions, hundreds of millions of authors that that administrative complexity of that is nightmarish. It could be done perhaps, but I wonder whether the transaction cost would be worth it, whether it could be done at all in fact.

Danielle Johnson: And Jane?

Jane Ginsburg: Okay. Article 4 may make it possible to create a market for licensing data. And and I agree with Bernt that the question is, "Who's going to be doing the licensing?" And I think that there may be two different universes at least. On the one hand, a lot of people, without knowing it, may have already licensed their works without knowing it, because they -- when they signed on to the terms of service of Instagram, et al., they may have granted a general use license to Instagram that is so broad that it would allow Instagram in turn to license AI entrepreneurs. So, it may be that as to many of those non-professional creators on the internet to whom Bernt was referring, it may be that perhaps unwittingly they already have granted permission to scrape their works. Then there is a different universe that one might call, quote, "high-quality"

inputs. And, as to those, I think there already is a certain amount of licensing going on. I believe recently the Associated Press has reached an agreement with -- I forget which AI entrepreneur. And there are probably good reasons for wanting to go to high-quality data, particularly high-quality data that can be or works that can be organized through an agency such as the Associated Press. Because, first of all, it makes it easier to license when you go to one place for lots of people's works. And the other is that, by licensing higher-quality data, it may be possible to reduce the so-called hallucinations that some of these AI outputs provide.

Danielle Johnson: Thank you. We'll take one last comment on this from Andres before switching to our next topic.

Andres Guadamuz: Just quickly in defense on the mundane providers of text, as one of such providers on Twitter, the models do not care about the quality of the content. They just care about the ingestion of large quantities of statistically significant text. There is a quality for some models that were trained on high-quality text. Reddit was an example in one of the early GPTs, I think GPT-2, was trained a lot on supposedly a good discussion from Reddit. So, a licensing system would have to consider that into -- consideration that -- take into consideration that for a lot of models. And I take Jane's point that potentially high quality may come into hallucination. There are people that are saying, by the way, that hallucination is never going to go away. That we're always going to have hallucination because this is just statistics and the machine doesn't care about the quality of the content. So, we are going to have to contend with this that the Mona Lisa has as much value as my scribbles on Twitter when it comes to data. So, that is one -- that is the challenge that I see with licensing.

Danielle Johnson: Great. Thank you. And I'll turn this back over to Chris.

Chris Weston: All right. Thanks. So, I want to thank everyone, first of all, for a great discussion on ingestion and training issues. For the remainder of our time, we'd like to focus on authorship and copyrightability. I'll kick us off with this question, which is, "What do international treaties tell us about machine authorship?" Does international law mandate particular approaches to authorship? Specifically, does it require human authorship? Jane, please go ahead.

Jane Ginsburg: I think human authorship is implicit in the Berne Convention. The Berne Convention does allow for non-human initial ownership in the case of cinematographic works. But, otherwise, one should not conflate initial ownership with authorship and the concept of originality. And I think there are many indicia in the Berne Convention that assume that the author will be a human being. Obviously, that text originates in 1886 when nobody was thinking otherwise. But the derogations from human ownership have not affected, I don't think, the concept of human authorship that is a foundation of the Berne Convention.

Chris Weston: Thank you. To anyone else? Andres.

Andres Guadamuz: All right. I think I'm probably going to be sticking my neck out on this one as the only member of the panel I think, as some are aware, that is going to dare to disagree with Jane on something. Apologies, Jane. I agree -- by the way, I completely agree with your point. I think human authorship is implied. Absolutely. That is, that has always been the case. My -- I'm going to make the argument, and I've been making it for over five years, that we need to relook at this because, and I think Peter mentioned this quite nicely in his presentation -- in the first presentation, is that we are going to have to start revisiting a lot of our human assumptions for lots of things in academia and in copyright, because the problem that we're going to encounter, and we are already encountering, is that we are not going to be able to tell if a machine has been used in the creation of a work. I know we're probably going to discuss this later. But, potentially, this is going to be a big issue. Already today -- just today, OpenAI announced that they were scraping their AI detector type of thing that they were going to be using. So, it is potentially problematic to -- for us to try to assume that no machine has taken place. The time in which we could tell when something had been created by human is soon going to be over. Now, what do we do with this knowledge? And, obviously, the UK has in Section 9, Paragraph 3 a provision. This was -- I am aware this was written entirely in a different time with very different type of computer-generated works in mind. It was more thinking about robotic painters that were completely following instructions from the programmers. But we may have to bite the bullet, if we want to call it like that, allow people who are potentially going to be using works that are going to have lots of AI or some AI or a little bit of AI to have some form of protection. And the UK provision, for example, is shorter. It's 50 years and it doesn't have any moral rights. So, maybe we can do something. And I know a lot of -- lots of the people in the panel are going to disagree on sui generis rights. I think lots of people do not like sui generis rights. But maybe something that is just some form of related right [inaudible] short term, it doesn't have to be 50 years, it could be 25 years. By the way, the UK provision has been in existence since 1988 and has never been litigated. So, the sky hasn't been falling. It's worth a consideration. I'm sorry, this is going to be probably my longest intervention. The UK IP Office actually conducted a consultation last year looking at Section 9, Paragraph 3 and decided not to do anything. They said, "Okay, everything is fine, everything -- we'll leave it. There is no litigation. We'll see how it continues developing." And they decided to keep it in the books. So, that's probably where we are.

Chris Weston: Okay. Thank you. Matt, I think, is next.

Matthew Sag: So, I think most jurisdictions have some concept of originality that seems to preclude purely machine-generated works. If you think that originality means creating something that expresses your own intellectual conception, and you can paraphrase that into different formulations for different jurisdictions, then works produced by a large language model, right, simply don't meet that criteria. Right? ChatGPT has no intellectual conception. It is guessing the next token time after time after time. You know, I think one of the major misunderstandings about these models is that because they produce things that if we had made them we would think were expressive and communicative, that the machine is expressing and communicating. And that's just not true. The machine has no internal mental state that it's trying to convey. So, when you look at the output of ChatGPT, it's text, it's not language. Right? Language involves that I have kind of a model in my head of what I want to say, and I have a model of you as the recipient, and I use words to try and impact you as the recipient. Right? The computer is just guessing the next word according to a very clever probability distribution. You know? And the image system is the same. But I do think we may need to be a little bit more liberal in our concept of what makes someone an author. I think that -- I don't think the Copyright Office is wrong in the "Zarya of the Dawn" case necessarily. But I think that maybe, you know, in some circumstances, you know, repeatedly versioning through a set of alternatives and then selecting and adopting one is something that we could start to understand as authorship. And, you know, I certainly agree we're all going to be using AI tools as part of our authorship. And, you know, to some extent, every time you write an email, you're doing something similar with predictive text. Right? So, I don't think, you know, the use of AI should be seen as a disqualification of authorship. You know, in a lot of cases, these things are just tools. But, you know, my answer to the actual question is, you know, are there, you know, international restrictions, you know, I see them inherent through concepts of originality.

Chris Weston: Okay. Thanks. We are coming up on the end of our time. I will ask Bernt and Jane to contribute. But then, unfortunately, that will have to conclude this discussion. I know that we all have more things to say in this very interesting topic. But, anyway, Bernt, please go ahead.

Bernt Hugenholtz: I fully agree with what Jane said about human authorship. This is ingrained in a system of copyright universally. So, we need human creativity. Can this be done with the aid of AI with very clever machines? The answer is, "yes." And, so, I agree with what Matt just said as well. We've just got to get used to the idea that creation can involve artificial intelligence. And raising the bar to the level that the U.S. Copyright Office has set in the Zarya case is perhaps asking a bit too much if you compare the high -- the very rigid standard there to the average snapshot, which is -- which will attract copyright all over the world, then there is a bit of a rift there. Having said that, there is still scope for -- obviously, for AI-generated output that will not pass that bar for that, there are neighboring rights. I don't think anyone has mentioned them yet, but there are international treaties that provide for neighboring rights protection. If you really want protection, you can look for protection under the phonogram right, under the broadcasters' right. Even in Europe, we have the database right. So, there are all kinds of IP regimes that protect non-creative output, which could be interesting for AI-assisted production that does not satisfy the copyright requirements.

Chris Weston: Okay. Thanks. And to conclude this discussion, Jane, please.

Jane Ginsburg: So, I would like to build on what both Matt and Bernt said. I agree that originality is the standard. And that it is the standard pretty much worldwide, even if it is not exactly the same worldwide. But I think there is a pervasive notion of human causation, as an element of originality. And that's where I'm not sure I agree with Bernt that the Copyright Office's approach in "Zarya" is too restrictive because I am sympathetic to the idea that reiterative commands that there's some level of creativity in reiterating the commands. But there's still the problem that the output remains something of a matter of chance. So, there is a lack of control. Now, does there come a point when the commands are so reiterative that the output could be attributed in a causal fashion? I'm not sure. But I'm sure the Copyright Office's ruminating on this as are other jurisdictions, because just selecting from a bunch of outputs which is the one you think reflects what you were trying to do, I'm not sure that there is enough there with that. Going back to what Matt said about the machine, of course, has no stamp of personality to put on the output. But we haven't talked about whether any attribution of authorship can be made to the people who program and train the machine. And it may be that the answer in most cases is "no," because they don't know what the downstream people are going to ask of the machine. But I am kind of curious to think about the extent to which there might be some form of authorship attribution to the upstream people. We seem to be mostly talking about the downstream people. And, finally, with respect to Bernt's comment about neighboring rights, I suppose the predicate question is, "If we can't find sufficient human authorship, why would we want to create a new IP right?"

Chris Weston: All right. Thank you very much, Jane. And thank you very much to the entire panel. We want to thank you all for participating in today's webinar. Video recordings of this webinar, as well as our other webinar on registration guidance and all four of our listening sessions, are available to the public on our website. The video from this webinar should be

available in three weeks. As Register Perlmutter mentioned at the start, the Office will be issuing a Notice of Inquiry calling for written comments on a number of questions on artificial intelligence and copyright. We look forward to hearing from you. Thank you, again, for joining us. And please have a wonderful rest of your day.