

Fordham Law School

## FLASH: The Fordham Law Archive of Scholarship and History

---

28th Annual Intellectual Property Law & Policy  
Conference (2021)

Fordham Intellectual Property Law Institute

---

4-9-2021 10:10 AM

### 7C Competition & Four Concurrent Sessions. Artificial Intelligence

Michael S. Shapiro

Heli Pihlajamaa

Janet Freilich

Carlo Scollo Lavizzari

Michael Fischer

*See next page for additional authors*

Follow this and additional works at: [https://ir.lawnet.fordham.edu/ipli\\_conf\\_28th\\_2021](https://ir.lawnet.fordham.edu/ipli_conf_28th_2021)



Part of the [Intellectual Property Law Commons](#)

---

---

**Authors**

Michael S. Shapiro, Heli Pihlajamaa, Janet Freilich, Carlo Scollo Lavizzari, Michael Fischer, Shlomit Yanisky-Ravid, and Ian C. Ballon

---

**Emily C. & John E. Hansen Intellectual Property Institute**

**TWENTY-EIGHTH ANNUAL CONFERENCE  
INTERNATIONAL INTELLECTUAL PROPERTY  
LAW & POLICY**

*Friday, April 9, 2021 – 10:10 a.m.*

**SESSION 7: COMPETITION & FOUR CONCURRENT  
SESSIONS**

***7C. Artificial Intelligence***

***Moderator:***

**Michael S. Shapiro**

*U.S. Patent and Trademark Office, Alexandria*

***Speakers:***

**Heli Pihlajamaa**

*European Patent Office (EPO), Munich  
Patent Law and AI – Out of the Black Box*

**Janet Freilich**

*Fordham University School of Law, New York  
Using Technology to Find Patents*

**Carlo Scollo Lavizzari**

*Lenz Caemmerer, Basel  
“Feed Me!” Said the Machine, and so Said Its Feeder: How to Square IP with  
AI, TDM and Machine Learning*

***Panelists:***

**Michael Fischer**

*Venner Shipley, LLP, Munich*

**Shlomit Yanisky-Ravid**

*ONO Academic Law School, Israel*

**Ian C. Ballon**

*Greenberg Traurig LLP, Palo Alto*

\* \* \*

MICHAEL SHAPIRO: My name is Michael Shapiro. I work with the U.S. Patent and Trademark Office, and I'll moderate this session. Like other moderators, Hugh asked me to introduce the subject matter of this panel very briefly. Our panel is about the intersection of AI and IP. In a nutshell, the continued growth of AI systems has challenged traditional notions of inventorship and authorship. At the same time, AI has given IP offices new tools to accomplish their work and achieve their missions.

My task has been made much easier by the panels that have preceded this one, in many ways beginning with yesterday's excellent panel on government leaders' perspectives on IP. The table has already been set with a banquet of issues on AI and IP. I do take note that the last panel treated AI as well. AI is everywhere.

Yesterday, Daren Tang, the Director General of WIPO<sup>1</sup> noted the interest of WIPO and the application of AI to intellectual property. I would note for the record that WIPO has been a leader in fostering discussion on AI and IP.

Also in his remarks, António Campinos, the Director of the European Patent Office, observed how AI is beginning to permeate a broad range of the EPO's activities, including by facilitating prior art searches, translations, research, citations, and classification of patent applications. Director Campinos also noted the importance of AI as a legal issue, pointing to recent EPO decisions, that an inventor designated in a European patent application has to be a human being and not a machine.

The USPTO<sup>2</sup> recently reached the same conclusion, rejecting two applications, reasoning, "The plain language of the patent laws as passed by Congress and as interpreted by the courts, limits patent applications to only naming natural persons as inventors." We'll have the opportunity on this panel to discuss the impact of those decisions on IP law practice and policy. What impact, if any, will they have on the pace of innovation? What changes in IP law will be needed as IP capabilities continue to increase? How will these decisions affect the fundamental goals of the patent system?

Shira Perlmutter, the Director of the U.S. Copyright Office also noted the critical importance of accurate and available data to the functioning of all copyright systems. She said, in particular, that the Copyright Office is analyzing the key issues related to the intersection of IP. Can a computer be an author? Are massive inputs of copyrighted works, for the purposes of machine learning, copyright infringements? Or should these activities be excused under an exception,<sup>3</sup> such as the fair use doctrine in the United States or the text and data mining exception in the European Union?

To discuss these and other leading issues, we have the good fortune in having a distinguished group of speakers to address these issues in greater depth and an outstanding group of panelists to continue the discussion. First, in her

---

<sup>1</sup> World Intellectual Property Organization.

<sup>2</sup> United States Patent and Trademark Office.

<sup>3</sup> This exception "allows researchers to make copies of works 'for text and data analysis'. This means that where a user has lawful access to a work they can make a copy of it for the purpose of carrying out a computational analysis of anything recorded in the work." Maurizio Borghi, Text & Data Mining, Copyright User, <https://www.copyrightuser.org/understand/exceptions/text-data-mining/>.

presentation entitled Patent Law and AI– “Out of the Black Box” Heli Pihlajamaa of the European Patent Office will explore some of the opportunities and challenges that AI presents to the EPO, including, as she stated in the program, myths surrounding patentability of AI-related inventions, the requirement of sufficient disclosure, patent eligibility and inventiveness, and principles of inventorship. Without further ado, I'd like to turn the floor over to Heli.

HELI PIHLAJAMAA: Thank you very much, Michael. Wonderful introduction from your side.

With that, I would like to start by submitting that AI is now everywhere around us. We have already seen that AI is something which is contributing to innovation, and which is really playing a very important role for the patent system. Playing an important role in order to manifest a patent system adaptable enough to also meet all the expectations that innovators and the general public have when they innovate with the help of or with AI.

One interesting aspect is that, even today, there is no clear definition or agreed definition of AI. What are we talking about? We have seen from the USPTO survey that there was a definition of AI. The World Intellectual Property Organization has a definition of AI. The European Union has a definition of AI. For the European Patent Office, the basis of our thinking is that AI is a broad sub-field of computer science that covers many different computational modes for solving data analysis problems empowered by aspects of human intelligence.

Normally, there is an aspect of data sets or rules set by algorithms defined by a human being. That is what we see at the moment in the patent applications of which we have seen a very rapid increase in the past decade.

What is important about patentability? Patent eligibility of AI is based in TRIPS<sup>4</sup> and other Conventions, and there it is stated that inventions are patentable in all fields of technology.<sup>5</sup> There, we again have a word which has not been defined. For the European patent system, just some weeks ago, we received a decision from our Enlarged Board of Appeal G1/19,<sup>6</sup> which explicitly said that “technology” or “technical” do not need to be defined at the moment because they need to remain open for further developments of technology, technique, and technical developments. That is exactly the point of why patent law can encompass all the new inventions and adapt itself to what is happening in technology and society.

The Enlarged Board of Appeal also stated that the so-called COMVIK approach based on a previous decision T641/00<sup>7</sup> is the basis of assessing patentability. Inventive step is the key because the first hurdle to have a patentable invention, which is a technical invention, is easily obtained. However, then only the technical features can be taken into account for the assessment of the inventive step, and therefore we always stay within the framework of technical inventions.

---

<sup>4</sup> The Agreement on Trade-Related Aspects of Intellectual Property Rights.

<sup>5</sup> Agreement on Trade-Related Aspects of Intellectual Property Rights, art. 22.1, Apr. 15, 1994, 1869 U.N.T.S. 299.

<sup>6</sup> Decision G 1/19, ECLI:EP:BA:2021:G000119.20210310 (10 March 2021).

<sup>7</sup> Decision T 641/00, ECLI:EP:BA:2002:T064100.20020926 (26 September 2002).

The teaching of the Enlarged Board of Appeal gives even further stability for the practice of AI related patents in Europe. What it also brings in is a further important aspect — I said this presentation is to show that AI patentability is out of the black box. That is because there used to be a lot of discussion about AI like "How do we know what is in there?" For patentability we need to know, because for patents, there needs to be made clear that any feature essential for carrying out the invention is described in the application in sufficient detail to render it apparent to a person skilled in the art. The person skilled in the art needs to know how to put the invention in.

That was stated for example, in a recent decision, T2574/16<sup>8</sup> by our boards of appeals, and we got another recent decision from our boards of appeals in T161/18<sup>9</sup> stating that it is important, for example, to describe what kind of training data is used and for which purpose that training data is used, so that a person skilled in the art can understand the invention. It is not necessary to provide the whole data set because that would overburden the patent system, but it is important to know what kind of data and how it is used to obtain the invention and fulfill the set criteria for sufficient disclosure.

Why is that also important? Of course, in the context of AI-related inventions, there has started to be discussion about bias, actually concerning ordre public and morality, which are established concepts of patent law. We don't grant patents for inventions where commercial exploitation would be against ordre public and morality. Therefore, we need enough details to disclose the invention in the application.

If we would be going to a deposit system, which might not even be necessary because all we're talking about can be presented in words. It might even be difficult to find the global agreement to agree on such a specific deposit system, which has been in the discussion — Such a deposit system would make it much more difficult to even look at that kind of ordre public and morality aspect, and the sufficient disclosure aspect. That would bring a further difficulty. How can a person skilled in the art get the information? How can a competitor get the information? How can society get the information?

Here I come to my last point, which concerns society, and the development of the societal understanding of all this that we at the forefront of technology are talking about. AI and the rights and obligations relating to AI need to be part of the societal agreement. If we would start what the USPTO wonderfully did not do, what the UK Patent Office did not do, what the EPO did not do, what the UK High Court did not do in their decisions, if we would start to grant rights to AI systems, how would that affect the later societal discussions about rights and obligations given to machines or computer programs, or how those are transferred somewhere?

Because, of course, the patent is a right to an invention. The right to an invention belongs to the inventor, as it is said in most of the patent conventions. If we would then say the inventor can be a machine, what kind of rights can a machine have? Can a machine have rights even going beyond the patent system? Not

---

<sup>8</sup> Decision T 2574/16, ECLI:EP:BA:2019:T257416.20191121 (21 November 2019).

<sup>9</sup> Decision T-161/18, ECLI:EU:T:2021:102.

forgetting, in everything that we are doing, the human-centric approach is important. Thank you for your attention.

**MICHAEL SHAPIRO:** Thank you so much, Heli. You've given us a great deal to think about, but we only have five minutes to discuss it. Let me quickly introduce our distinguished panelists. We have Michael Fischer from the Venner Shipley firm in Munich; Shlomit Yanisky-Ravid, from the ONO Academic Law School in Israel, a Visiting Professor at Fordham Law School (from 2012) and a Professor Fellow at Yale Law School, ISP;<sup>10</sup> and Ian Ballon from Greenberg Traurig in Palo Alto. They all have questions, no doubt, but let me take the prerogative of the moderator of perhaps posing the first one. At a most basic level, should an AI system be recognized as an inventor?

**HELI PIHLAJAMAA:** As I said, inventorship is a right. In our view at the EPO, and my view, and as I have seen from the court decisions of the other jurisdictions as well, to give that kind of right to a machine might be, at the moment, a little bit premature. That would require a lot of further assessment and further assessment in relation to other systems of rights and principles of law.

**MICHAEL SHAPIRO:** Thank you for reaffirming and confirming that basic point. Other panelists, feel free to jump in.

**SHLOMIT YANISKY-RAVID:** If I may, thank you for inviting me, and thank you for this interesting discussion and conference. What I would like to say is something we wrote in the article that will soon be published in the Michigan State Law Review about summoning the new artificial intelligence as a patent model. We raised these issues — part of them, you've discussed very beautifully, like the black box, and you gave some answers — but we still have some issues about the person of ordinary skill of the art and a machine which can view numerous amounts of data, and also inventors giving false details because firms and the startups don't want to say, "We just use AI." We gave examples of AI as creative inventors.

There is also a matter of quantity because AI can produce — like we show with electronic cycles — a huge number of patents and inventions and can submit them by itself. I just want to ask you: What do you think? Because we raised a new model, my co-author, Regina Jin, who is a patent officer, and I suggested having an AI patent special track to deal with all these AI issues, rather than just thinking about them and testing them as part of the regular track. What do you think about that?

**HELI PIHLAJAMAA:** Wonderful questions. Exactly what is part of the key issue here. Of course, in my seven minutes, which I overdid, I couldn't touch upon all those parts. Certainly, the AI creating new inventions, the AI creating new prior art, that might be a fact, that is certainly a fact, but of course, on the other hand, as Michael mentioned at the beginning, AI is also giving the patent offices the tools to search the data and to classify the data, and thus overcome, let's say, the huge crowd in the prior art in the same way.

On the other hand, the person skilled in the art — so there we can see like we have seen in many other technical areas which have been developing — for example, in chemistry, in some areas, it's very clear that the person skilled in the

---

<sup>10</sup> Information Society Project.

art is expected to know how to use a centrifuge. They might not even themselves always be able to know what happens in the centrifuge when they put some chemical substance in there. In the same manner, it can, of course, be expected that the person skilled in the art knows how to use the basic AI tools, and how to employ AI in the processes.

That is part of the adaptability of the patent system because in all jurisdictions, the concept of a person skilled in the art is not a fixed one. It is always having a different level, with a little bit of focus on what is happening in that technical area. Of course, there is one point about what you said also about the disclosure. Namely, we saw in the first AI applications, we started to receive a debut.

There, it often happened that the applicant just brought in “doing this using AI” or “achieving the invention with the help of a neural network,” to put it in very easy ways. Of course, that is not sufficient disclosure, as we have now received the feedback from the Board of Appeal decisions which I mentioned in my speech.

MICHAEL SHAPIRO: I see that the clock has run on this particular round. Unless I hear a really compelling need to pose a question, which could be picked up later by Heli, I'd like to turn now to Janet Freilich of the Fordham University Law School, which I understand is one of the premier institutions in the world. She will address the intriguing topic of using technology to find patents.

JANET FREILICH: Thank you so much. I will confirm that Fordham is indeed one of the premier institutions in this area. Thank you to Hugh for hosting this conference. Let me talk about how technology is changing the way that we search for patents, find patents, interact with patents, and then some of the uses and pitfalls of technology.

Historically, we just ignored a lot of patents. It used to be that there were so many patents out there, that it was impossible for any one person to read all of them or be aware of all of them. Computers, of course, are very good at reading large numbers of technology and aggregating that information. There is a market that has popped up over the past few years, called patent analytics as an umbrella term that uses computers' AI technology to track information in patents. The market right now is set with approximately \$700 million per year. That's projected to grow to \$1.7 billion a year in five years. It is a growing technological area and it's of growing importance.

I'm going to focus here, since I only have a little bit of time, on automated freedom to operate analyses, and on automated patent landscape searches. I will discuss the uses and the challenges of both of those technologies. Let me begin with automated freedom to operate analyses. Here, what's happening is AI is being used to conduct a freedom to operate analysis on the particular technology. Probably not surprisingly, this is most advanced in chemistry because it's a lot easier to define claim scope in chemistry than in other fields. Although, I am not saying that this is an easy task even in other fields. Because it is challenging, automated freedom-to-operate analyses are not in widespread use right now. They do have a long way to go before they will be in widespread use, if ever.

But there is currently at least one service that allows scientists who are thinking about synthesizing a particular molecule to design a method of synthesis



that avoids patented methods or patented molecules, which sounds very nice in theory, because if the computer can do your freedom to operate search, first of all, it might be possible where it was not before, it's cheaper, and in some circumstances, it might be more reliable than doing a search manually.

There were also some major challenges. When I talk to scientists about automated freedom to operate searches, they will say, "Oh, wow, that sounds great." Then when I talk to lawyers, they say, "Oh no, that sounds very dangerous." One of the major challenges is that right now, automated freedom to operate searches effectively weigh all patents equally. So, if you find a group of patents that is relevant, the computer will say, "Avoid them all."

That is not necessarily true in practice. In practice, some patents that, at first glance, appear to pose a roadblock to introducing a technology might turn out, first of all, not to be valid — many patents are not, in fact, valid — or claim language can be ambiguous. Patents may not actually read on to the technology issue if the claims are closely analyzed.

All of these concerns can be dealt with, with manual review, and perhaps with AI as it improves. Right now, AI is not particularly good at assessing validity and not particularly good at interpreting claim language.

One of the challenges of automated freedom to operate searches is that they make it very easy to bypass lawyers. If you are a scientist, you can save money on a freedom to operate search instead of asking a lawyer to do it, by inputting your desired end result into the computer, and then your computer spits out some recommended synthesis route, for example.

But there is a danger to not seeing the patents you're avoiding and to not having a legal expert analyze which patents you're avoiding. My main concern is that automated freedom to operate analyses might lead to over-avoidance of patents and might increase the cost of designing around a patent because it requires designing around patents that may not actually need to be designed around and may over-deter some new technological development.

I also want to talk briefly about a second type of patent analytics, so not freedom to operate searches now, but patent landscape analysis, which is a broader category of analysis that uses AI to review large numbers of patents in a particular technology and produces a report on the patent landscape. The report might include details, it may tell how many patents exist, who owns the patents, what subcategories of technology these patents relate to, and other details as specified by the user.

The idea is that policymakers and businesses can quickly survey the state of technology in the field and make decisions based on that knowledge, which is — it's very exciting to be able to do this with patents. It's also something that has been done for a long time. Technology is increasing our ability to search, classify, organize, and extract information. It's making the process cheaper, and it's reducing the amount of manual review that needs to be done.

There are policymakers who have commissioned patent landscape analyses, the WHO and the UN have produced a number of patent landscape reports. National governments have used this technology to understand collaboration and knowledge flows between companies to select areas that tend to focus to devise innovation

policy. Companies will use patent analytics to get information on competitors, to make decisions about moving into new markets, to inform merger targets, possibly to do what are called whitespace analyses, to highlight where patents are not prevalent, which provides companies with leads on underdeveloped technology spaces they might want to move into. That is a really exciting possible use of the information in patents, which has historically been hard to extract because there's just so much of it.

There were also pitfalls to automating this kind of patent landscape analysis. Like I described for freedom to operate analyses, patent landscape analyses often weigh all patents equally. They do not necessarily account for validity. If you have two patents in the same technology area, those two patents do not necessarily reflect equal amounts of research. In fact, they often don't. One patent might reflect a lot of research in a core area, and another one might reflect something that has never actually been tried.

My concern is that patent landscape analysis can be really misleading if it's interpreted by stakeholders who are not familiar with patent law. It can make it look like there's a cluttered patent landscape, where that actually might not be true. There might not be valid patents or the patents in the area might deal more incidentally, and not reflect the actual research in a particular area.

Let me end just with a couple of cautions about this technology. First of all, I think that patents are going to become a lot more accessible. I'm totally out of time. I will just stop there and say this is interesting technology, but there are dangers. Thank you.

MICHAEL SHAPIRO: Thank you so much, Janet. It was a very clear statement accessible to even a person not schooled in the technical arts, such as your moderator. I would just turn to other members of the panel to see if there are any questions at this time.

HELI PIHLAJAMAA: Janet, if I may. I found it fascinating what you were saying, especially that AI can make patents and technical knowledge in patents more accessible. Could you imagine how the understanding problems could be solved in that context as well? Because you pointed to that danger.

JANET FREILICH: I think one key point is simply to always have a lawyer help look at this. AI does not mean getting rid of lawyers. I think it's just really important to remember the role of lawyers in interpreting these. I also think it's important to have lawyers on the design side of these technologies.

There are many different services that offer these. I'm deliberately not naming specific companies. Some companies seem to be mostly driven by computer scientists and others by lawyers. It's good to have a combination because both areas of expertise are really important.

It's very easy to get information from patents, but if your program doesn't know what the information means, it can be very misleading. I would say both on the design and on the interpretation, just make sure to talk to patent lawyers. Thank you for your question.

CARLO LAVIZZARI: Janet, would you say there is a particular demand in chemistry, especially due to the way the patent system operates in chemistry? Or do you think it would be applicable in all fields per se?

JANET FREILICH: I think that if a technology were available there would be demand in all fields. It's just much harder to do as well in other fields. To the question on demand, it's an interesting question because at least in the US, there are reasons to not want to know what patents are out there when you are developing a product. You might have increased damages if you do eventually infringe, if you were already aware of those patents. At least, anecdotally sometimes engineers are told not to read patents, to avoid knowing about it.

To the extent it becomes easier to read patents or easier to know about patents, I'm not sure that scientists would necessarily see that as a positive because it also potentially increases the dangers of willful infringement. That said, if it's easier to sort through patents, it could be helpful in some industries. It could also make everything grind to a halt. There are some industries where it would be impossible to do anything if you actually took into account every patent that was out there.

As a practical matter, a really good freedom to operate search could actually be a really problematic thing as our patent system stands right now. I think patent landscape analyses are more promising and could be useful in all industries.

MICHAEL SHAPIRO: Well, good questions all and terrific responses. Are there more questions at this point before moving on to our final presentation? We'll also have an opportunity in the wrap-up general discussion, if we've missed any.

SHLOMIT YANISKY-RAVID: I just wanted to mention some issues that I thought were really interesting. I think heading into this era where there is AI for doing searches, transparency is really important — for the purpose of giving the client the choice whether he or she would like to adhere to an AI search or not. Alternatively, if there is a gap between the lawyer's decision and the AI, "someone" has to explain why. I think in the long run, heading into the future, I like what you're saying, that AI may save a lot of money and these systems are very efficient. I think, at a certain point in the near future, it would be a negligence *not* to use AI or not to offer AI. I really think sooner than you think, the use of these AI systems will be implacable and feasible. Thank you.

JANET FREILICH: Thank you so much. You're 100% right about transparency and I just wrote that down in big letters because that's going on my paper. Yes, transparency about what the AI is doing: very important. Thank you.

MICHAEL SHAPIRO: Well, good. I'm learning quite a bit. I think it's now time to make the transition from the patent space to the copyright space. To help us in that new territory, we turn to Carlo Lavizzari, who has a presentation titled, "'Feed me!' Said the Machine, and so Said Its Feeder: How to Square IP with AI, TDM, and Machine Learning." Carlo, the floor is yours.

CARLO LAVIZZARI: Thank you very much. This is borrowed in fact, from the *Little Shop of Horrors*, for those who know the musical where a carnivorous plant ultimately has to be fed to keep going. The entire venture of a flower shop that just wanted to become a little bit more famous ends pretty badly for at least the creators involved. Some of the characteristics of AI, which heavily is built on one side may be patents, but also trade secrets and what patent lawyers call nonpatent literature: scientific, technical, medical, and other literature to derive

the knowledge to, in fact, teach the machine, analogous to feeding the machine, to create value for humanity.

In essence, we have to marry with the new technology of artificial intelligence what creativity offers — the art of the impossible — with a political sphere, which is the art of the possible, as was said by Ben Okri, a writer.

At the moment, I think most IP lawyers as I can see from AIPPI<sup>11</sup> gravitate to being slow to adapt the law and apply existing norms. I think I would subscribe to that too. Similar to what Heli alluded to, I think it is important to take a light step back and say, “What are the policy frameworks for humanity for a human-centered AI approach that are needed?” How does intellectual property and, in my case, copyright fit into it?

I think that is important because not every AI may be, in fact, beneficial, a bit like the carnivorous plant. It could come with some dangerous or unwanted effect or just become a runaway technology. I think when you look at genetics, the bioethics field with many professors with tenure at universities investing time in the ethics of genetics and research in that area, maybe similar ethics are needed for the field of AI as well.

Apart from personal data, freedom of information laws, equality and de-biasing, the reuse of data collected with public money or private money, the specific concerns with health data and data sets, the transparency needed for statistical decision making may be important factors. If you think of HR, people being either shortlisted for jobs or not, and then perhaps being happy in the future— for these decisions to be taken entirely by machines as long as they have been preset in a defensible manner.

One could break this down to transparency and accountability. Quality a bit like in the text and data mining discussions, the good old rule of garbage in, garbage out. It is paramount to have quality there. The privacy and security issues of consumers and those whose data might be used, the fairness availability issue and then sustainability or architected-in values that may come up.

Just two or three days ago in Europe, a cultural committee of parliament in Europe said that cultural diversity should be built and regulated into AI. There could be a number of value factors coming into play. With this macro view, which I mention because I do think even in the limited seven minutes of time available, it's important to keep this larger perspective, my thesis is that, in fact, protecting IP used as inputs for AI is conducive to the values of building trust, high quality inputs, transparency, and assuring the provenance and guarding against — let's say, fake news being used as input to a good AI technology which will still produce terrible outcomes because the fundamental basis is either untraceable or unknowable. My thesis is that what is needed is a sustainable environment where the machine wants to be fed but so do the creators.

Already for the STM<sup>12</sup> field, a vibrant licensing network in fact exists for machine learning. Text and data mining has really come a long way since when this technology started to appear about 40 years ago. For the publishers that want to be involved, they can either use AI to become more efficient or offer trusted data sets

---

<sup>11</sup> International Association for the Protection of Intellectual Property.

<sup>12</sup> Science, technology and medicine.

that have a certain quality for input and offer those for sale. You would say that the future reader for authors and for publishers perhaps in the nonfiction field, perhaps even in fiction for sentiment analysis, et cetera, are machines. So, the AI space not only changes who the inventor might be, but also who the consumer might be.

There's another important report from the EU parliament about education and how AI will change the entire educational landscape. For educational publishers, surely if the consumer environment changes, the learners learn in different ways. That changes the entire market as well.

When we look at the existing landscape of existing IP laws we have, through fair use and exceptions that are perhaps widely worded, a vagueness which many welcome, but I would argue is detrimental, leading to quality inputs being under-resourced and under-demanded, and an obscurity — how content is actually obtained. I think in many cases, the trade secret IP protection desire leads also to data sets used for machine learning to remain obscure, perhaps the providence even being illegal or questionable in other ways.

I think for the entire trust environment and in an ecosystem of transparency — where does the training data come from? Can I trace it? Similar to drug experiments, where in a trial I have to be able to disclose every syringe I've used, I think here the transparency would do a great deal also to foster consumer confidence. Licensing, I would argue, is in fact conducive to this integrity and transparency.

I come now to a discussion between licensing and exceptions. I think many of the big tech or promising AI proponents would argue that exceptions are needed so inventions can make good use of existing IP data sets — that the users are non-consumptive and that there is no interference with the market or that whatever is extracted are just mere facts. Those would be typical arguments. I think yesterday the *Google v. Oracle*<sup>13</sup> software case was mentioned, which may fuel this line of arguing.

I would hold against the notion that fair use can also entrench deep-pocketed commercial users. It is not always in the interest of leveling the playing field to have fair use apply. As to the non-consumptive views, I would say if there was a whiskey distillery and it found that the whiskey it produces is used for biofuel in a non-consumptive way, it would still want to be paid for its whiskey. Of course, in the case of whiskey, it would be a sacrilege to use that for biofuel.

A case that has gone the other way is certainly the *TVEyes* case<sup>14</sup> in the US that shows that AI is also subject to copyright rules. One could also cite the *Audible*

---

<sup>13</sup> *Google LLC v. Oracle America, Inc.*, 141 S.Ct. 1183 (2021).

<sup>14</sup> *Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169 (2d Cir. 2018).

settlement<sup>15</sup> in part that AAP<sup>16</sup> achieved. More recently, in a lower case<sup>17</sup> from the United States District Court of Delaware just a few days ago, my attention was drawn by Roy Kaufman, where it was held in a dispute that the case can go ahead — a case involving analytics of legal text. The plaintiffs were Thomson Reuters and Westlaw against an entity called Ross, Inc. that wanted the case to be dismissed. The Delaware court found, “No,” there was a sufficient averment of copyright violation and also tortious interference with contract, which is perhaps another avenue for rights holders to protect their AI — having a stringent contractual chain.

Having rushed through my thesis of saying licensing is conducive to building trust, I want to also recommend next steps in this open field, and I'm happy to be shot down on them. I think we would need some standard identifiers for AI entities, whether they are recognized as legal entities or not — I would favor not. Still, identifying and giving an identifier to AI would make the transparency easier and would be basically faster and perhaps even a deposit system.

The provenance of integrity could be built along the lines of escrow agreements. There is a WIPO Treaty on the deposit of microorganisms. Maybe one could look at some kind of deposit rule for data sets that have been machine trained and rules for disclosure that limit the disclosure to appropriate cases, safeguarding trade secrets, and legitimate expectations.

Finally, I would say that a whole cultural debate needs to take place — an ethical debate similar to the bioethical debate in genetics. Of course, I would argue that we ought to have fun and debate in such conferences, such as the one Fordham has a long tradition in. That would conclude my talk today.

MICHAEL SHAPIRO: Carlo, thank you so much for those insightful remarks as always. We have a number of questions, many of them focusing on licensing issues that you've already addressed. I also want to flag that Ian on our panel has some observations on data portability including copyright, the DMCA,<sup>18</sup> and non-IP protections that we would like to save some time for.

Let's begin with what I will call a group of questions relating to the licensing issue. One commentator put it this way: Companies which use copyrighted works to train AI — ought they be paying licensing fees for the reproduction right? I think we can anticipate the answer to that, but I'll let Carlo speak to it.

Another questioner asked: Is collective licensing useful? It may be critical for efficient permission seeking and granting for machine learning uses, but should collective licensing be required? Let me just leave that as a paired set for Carlo's response, and then we'll circle back to Ian for some views on data portability.

---

<sup>15</sup> Audible, an audio book business, agreed not to include the copyrighted works of several authors in one of their programs without receiving the consent of said authors. Andrew Albanese, In Captions Settlement, Audible Will Not Use AAP Member Content Without Permission, Publishers Weekly, <https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/82370-in-final-settlement-audible-agrees-to-limit-its-captions-program.html>.

<sup>16</sup> Association of American Publishers.

<sup>17</sup> Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc., No. 20-613 (D. Del. Mar. 29, 2021).

<sup>18</sup> Digital Millennium Copyright Act.

CARLO LAVIZZARI: Yes, thank you for those questions. I think the answer is actually, it depends, to that one asked by Barry Scannell. If you look at the Google Books case,<sup>19</sup> if all you want to do is count the number of A's or the genetic codes in a science article, just the genes as they are mentioned, perhaps these uses are indeed fair use and do not really require much.

Interestingly, in the *Google* case, which is often cited for not needing to pay, the judge in the second circuit excluded look-up information, cookbooks, poetry, short works that would be damaged in very different ways from longer works where it was found fair use to copy entire works to then do something useful, for instance, the visually impaired.

I think the answer is, it depends, but there isn't a one-size-fits-all answer. It really depends. I would argue that for most AI-sophisticated uses, and that can perhaps already answer the second question, when you have, for instance, 16,000 diabetes articles you want to analyze for a decision tool for a doctor, they're not all coming from the same publisher. They might be coming from a finite number of publishers, in which case individual licensing is possible, but in many instances, it will not be possible to achieve a literature that is comprehensive and high quality and doesn't miss out.

I think in the case of diabetes, we would all agree, you can't afford to not want to know, as perhaps people might say in the freedom to operate scenario. There, perhaps collective licensing is the method of choice. I'm not sure it needs to be mandatory, per se, but perhaps if no easy licensing is available, that could be a factor, pointing again to some kind of fair use.

MICHAEL SHAPIRO: Thank you for that, Carlo. I should mention that the question on collective licensing was presented by Richard Ruddik. I'm not seeing additional questions, so perhaps this is a good moment for Ian to present some observations on data portability, if you're available to do that.

IAN BALLON: Yes, thank you. Let me also provide a brief overview of US protection. Carlo was talking about licensing of copyrighted material, and indeed that's part of the framework, but in the United States, an overlay that we have is not just fair use, but also the First Amendment. The legal regime in the U.S. is quite different from Europe, and in practice, most companies that are using data sets in the United States are dealing with unprotectable data.

Certainly, if you're dealing with the case that Carlo talked about — protectable articles or other copyrighted materials in a database — then there's actually a split copyright in the United States (under section 201(c) of the 1976 Copyright Act).<sup>20</sup> For literally 19 years, I represented Knight-Ridder, and then McClatchy in *In re Literary Works in Electronic Databases Copyright Litigation*,<sup>21</sup> which went to the U.S. Supreme Court — on a narrower issue — as *Reed Elsevier v. Muchnick*.<sup>22</sup> That case involved the split copyright of 17 U.S.C. § 201(c) where both the database owner and

---

<sup>19</sup> Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

<sup>20</sup> 17 U.S.C. § 201(c).

<sup>21</sup> *In re Literary Works in Electronic Databases Copyright Litigation*, 654 F.3d 242 (2d Cir. 2011).

<sup>22</sup> *Reed Elsevier, Inc. v. Muchnick*, 559 U.S. 154 (2010).

contributors own copyright interests. If a data set is comprised of copyrighted material from a database, there may be fair use issues. If the material is used to train an algorithm, it would likely involve intermediate copying. Unless the scraping is done in real time, typically there's an intermediate copy that's downloaded, then data is extracted, and the downloaded copy of the work is deleted. The question then would be, "Is that a fair use?" The answer would depend in part on the impact of the market for the genuine product. There are examples of commercially licensed data sets. But there are also examples where there is no market that is adversely affected when data is used for machine learning and indeed the use itself may be highly transformative. *Authors Guild v. HathiTrust*, is an example of that.<sup>23</sup>

But using copyrighted material to train AI is not the typical case in the United States. Fundamentally, when we are talking about data inputs — the vast majority of test sets comprise data that, under *Feist*, is not entitled to copyright protection.<sup>24</sup>

Database owners have a range of remedies, and companies looking to use data have a number of potential legal obstacles, copyright being one of the most significant because of the potential availability of statutory damages and attorneys' fees, but on the other hand factual databases are entitled to, at most, only thin copyright protection. If an entire factual database is copied, that will likely be infringing — assuming creativity in the selection, arrangement or organization of the database. Courts apply the substantial similarity or virtual identity test — or in the Second Circuit, "enhanced substantial similarity" — but typically when you're dealing with data sets, only specific elements (and not an entire database) are extracted.

The other strong remedy for database owners is the anti-circumvention provisions of the DMCA, to the extent that data is extracted in a way that circumvents either technological measures or password protections. But there is actually a range of other (non-IP) remedies that typically come into play, including certain common law claims.

MICHAEL SHAPIRO: Ian? I'm just a little concerned that we're bordering on a separate presentation, so if you could be concise on this so that we can provide opportunity for other questions?

IAN BALLON: Sure. I would just say that there is a smorgasbord of remedies depending on the manner in which data is extracted and used, but much of the data in the United States is actually factual, and unprotected under copyright law. That's a far different regime than what exists (or may exist in the future) in Europe.

MICHAEL SHAPIRO: Well, thanks so much for that. I do appreciate it, and sorry to interrupt, but I think you were able to make your central points.

Are there further questions? [Unintelligible] poses a question: Article 3 of the DSM<sup>25</sup> directive addressing cultural heritage institutions — what is the lawful access? Article 4 of the same directive also addresses lawful access and rights

---

<sup>23</sup> *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

<sup>24</sup> *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991).

<sup>25</sup> Digital Single Market.



holders, including a unilateral opt-out provision. He asks Carlo whether he has any comments on these articles.

CARLO LAVIZZARI: On the non-commercial uses for heritage institutions like libraries, museums or cultural institutions, they are benefiting from an exception and can do text and data mining over broad sets that they typically already hold through prior agreements, be that subscriptions or individual legitimate access arrangements with typically publishers and intermediaries. To the extent that they have lawful access pre-purchased, they can engage in text and data mining, which is presumed already accounted for. They also license very large, broad sets across disciplines.

Meanwhile, commercial companies are able to text and data mine for commercial purposes unless publishers, similar to what Ian Ballon said, where publishers or rights holders are commercializing the selling of sets for text and data mining, they can opt out and declare so. The practicalities of how these opt-outs are being recognized are being worked out in Europe as we speak.

MICHAEL SHAPIRO: Thank you for that response, Carlo, and thank you for the question.

SHLOMIT YANISKY-RAVID: I agree with both of you that the data is more important than even the software. As Ian said, it's not protectable within the US. However, I argue that this issue raises a different kind of protection of a trade secret, because I look at some core decisions such as the *Loomis*<sup>26</sup> and *Robbins* where courts held the opinion that the data that the AI system was based on is a trade secret and therefore, it cannot be exposed. I think this decision raises a very interesting question, such as, what is the interaction between IP and trade secrets, and what can be done about that? Data is surely very important and more so essential for AI systems in order to operate efficiently.

MICHAEL SHAPIRO: Ian, did you have any response to that, or any other members of the panel?

IAN BALLON: I don't have a specific response to that. I will say, because I know we're short on time, I did include a chapter from my treatise on the U.S. perspective on database protection and data portability in connection with AI, for those who are interested in the materials.

CARLO LAVIZZARI: I would have a response. That's precisely why I mentioned the idea of escrow agreements or some kind of deposit like the treaty on microbiology organisms where, under certain conditions, you can access the data if you have some kind of legitimate reasons. If we are creating machines that use data and the data that's used for learning is simply protected by trade secret, it becomes completely opaque as to how these machines operate.

HELI PIHLAJAMAA: From the patent law perspective, I think that the disclosure requirement, as I described it, at least in Europe according to the European case law there, would make it necessary that there is enough data, not necessarily the data sets, but enough information about the purpose and the scope of the data in the patent application.

---

<sup>26</sup> State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

The Budapest Treaty for deposit of microorganisms<sup>27</sup> — there, I see some difficulties when we are talking about data and the patent system. Of course, for the microorganism, the necessity was there because you cannot put the microorganisms in paper. You need to have a place where they are fed, as Carlo was feeding the machine earlier, so that they can survive for the lifetime of the patent. Therefore, you need a specific rule — who can have access and how, and whatever expert solutions are there.

There we are talking about totally different volumes than what we are talking about if we would start to have such a system for AI patents. Also, it would have to be a global system. That is a challenge, as I mentioned before. Of course, for other areas as trade secrets or data protection purposes, like Carlo and Ian mentioned, there might be some other openings.

IAN BALLON: I think it would inhibit innovation if we had a permission-based system. It's really not practical. AI works based on test sets, and you have to feed the machine with data to allow for machine learning. There are billions and billions and billions of inputs every single day being gobbled up for the progress of the useful arts and sciences.

If we bring that to a halt and put in place a permission-based system — and only “feed the machine” those inputs that can be licensed — AI won't work very effectively. You can't partially train an algorithm and expect it to function properly and be able to simulate artificial intelligence. There are, certainly, big data owners who are licensing data sets. I think licensed data sets can coexist, but we can't have a permission-based system only or it will kill the effectiveness of the technology.

MICHAEL SHAPIRO: Quite a useful exchange. Thanks both for the question and the response on that. I see that we now are indeed seven minutes over our allotted time of one hour. I'm sure we could go on for another hour, but I don't think that's permitted under Hugh's rules. Of course, when we're back in New York next year, we'll be able to continue these conversations over meals and glasses of wine. I know we all look forward to that.

Please hold on to those questions. AI is not going away. It's everywhere. It's everywhere. No doubt we'll have an opportunity next year to continue this conversation. If there's not any other additional comments or questions or last thoughts, perhaps we could call it a day for now.

\* \* \*

The below question and answer were submitted after the conclusion of the session:

MICHAEL FISCHER: Can you foresee the impact of the decision G1/19 on the patentability of AI inventions before the EPO?

HELI PIHLAJAMAA: Since the decision G 1/19 set a clear line for the assessment of the patentability of any CII, including AI, the decision clarifies the assessment of patentability of AI-related invention at the EPO. The principles of

---

<sup>27</sup> The Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure, Apr. 28, 1977, 32 U.S.T. 1241.

the COMVIK-approach are applied. The disclosure and claim drafting will need to allow the application of the COMVIK-approach. All in all, the decision G 1/19 gives further stability for the patentability of CII and AI at the European Patent Office.