Center for Strategic and International Studies

TRANSCRIPT
Event
LeadershIP 2024
**(Panel II) Intellectual Property and Artificial
Intelligence**

DATE
Tuesday, April 9, 2024 at 11:00 a.m. ET

FEATURING
Lynne E. Parker
Associate Vice Chancellor and Director, AI Tennessee Initiative, University of Tennessee

Rama Elluru
Senior Director for Society and Intellectual Property, Special Competitive Studies Project

Phil Hill
Special Counsel, Covington

CSIS EXPERTS
Kirti Gupta
*Senior Advisor (Non-resident), Renewing American Innovation Project, CSIS: Co-Founder and
Executive Director, LeadershIP*

CSIS | CENTER FOR STRATEGIC &
INTERNATIONAL STUDIES

Kirti Gupta: All right. Welcome back, everyone. This is our panel at the intersection of artificial intelligence and intellectual property. And as you all know, you know, artificial intelligence has been around for some time, but really the world changed on November 30th of 2022 when OpenAI introduced ChatGPT to the world, and the rest is history. It had a hundred million users within a month. It's one of the most fastest-adopted technology platforms. And that really brought generative AI at the forefront.

And a lot of the issues that are – that we are seeing at the intersection of artificial intelligence and patent/copyright AI issues, IP issues, are focused around gen AI. So, you know, we are going to be focusing our conversation on those. And it's really important to be talking about this technology and other emerging technologies like we have been because, you know, we are at CSIS. We are talking about technology leadership in critical emerging technologies. And nothing more important than AI, which has changed our world, leveled the playing field for defense, national security, and other technology leadership purposes.

So, with that, I'm really thrilled to introduce this perfect panel of experts to be exploring this. And I'll say why in a minute as I introduce all of them.

So on my right is Dr. Lynne Parker. She is associate vice chancellor at the University of Tennessee and the director of the AI Tennessee Initiative. And prior to this role, she led the national AI policy efforts from 2018 to 2022 in the White House Office of Science and Technology Policy, serving as the deputy technology officer of the U.S. and founding director of the National AI Initiative Office. She's been at the UT faculty since 2002. So when we talk about the technology issues, I'm really happy to have Dr. Parker on our panel with us.

And then we have my old friend Rama – (laughs) – Elluru sitting next to her. She's a senior director at the Special Competitive Studies Project, SCSP. It's a nonprofit based here in D.C. established to make recommendations to strengthen America's long-term competitiveness for future of AI and other emerging technologies that reshape our national security and economy. And before SCSP, she was the director at the National Security Commission on AI, which is when we met, I think. And before that, she was an administrative patent judge on the Patent Trial and Appeal Board at the USPTO.

And then we have Phil Hill, who is a special counsel at Covington. And he focuses on complex copyright matters, with an emphasis on music, film, TV, videogames, sports, technology. And he's been representing clients in federal and state court and before U.S. Copyright Royalty Board, Copyright, Patent, and Trademark Office, et cetera. So he's been regularly advising clients on cutting-edge topics like gen AI, Metaverse, NFTs. So he can give us a really good landscape on the current litigation in the patent/copyright space in AI.

So with that, I think we would jump straight into the conversation in this panel, and then we'll invite participation from all of you. So I thought it's the – one way we can kick this off, because we have Dr. Parker on our panel with us, is to ask you the first basic question: Why here? Why now? Why is the AI revolution here today? Why is it here now? And where are we headed?

Lynne E. Parker: Well, I think most of the people in the audience are aware that AI, of course, is not new. It didn't just show up about two or three years ago, but it has been around since the '50s. I think what is different now is that we've digitized the world. We've also created computers that are extraordinarily powerful and also at our fingertips, and so many organizations can have access to that computational power. And the algorithms that have been around for decades have now been tailored to that situation where we have lots of data digitized about almost every aspect of the world and we have that computation that's broadly available.

So what does that mean? So I started working in the White House about – it was August of 2018, and at that time there was – it was extraordinarily rare for the term "artificial intelligence" to ever show up, say, on the national media, on a news station for instance. It would almost never happen. And then in no time it started growing, and so that was pre-generative AI in the – in the sense of, you know, these tools now being freely available. And so what was beginning to happen around 2018, 2019, 2020 is that the digitized world, with those algorithms and the computation, the capabilities of AI systems to do things like recognition and language processing and so many other kinds of task(s) was beginning to exceed the capability of people. This was true even before ChatGPT was released. And so, in the sense that it's surpassing the ability of people to do things like image recognition, language processing, and so forth, that meant that industry was beginning to adopt it. And so we saw that adoption happening around 2018, 2019, 2020.

And that was when people began to recognize some – not only the power of AI to very much create new capabilities, address important challenges across the board, but also some of the early potential harms were beginning to be recognized. If you go back to around 2018, for instance, there was a famous example where Amazon was using AI to help screen resumes because they had so many and they wanted to find the ideal candidates to come and work at Amazon. And so they used AI. Well, lo and behold, the AI detected a pattern in the data, which was that, well, apparently, to be successful in a technology company you need to be male, because most were. And so, therefore, it screened out all the women and just presented, you know, mostly the male resumes. Well, it's not because the AI was designed to be intentionally biased, but it was perpetuating a pattern that already existed in the data, which is that most people that work in technology companies apparently are male.

And so that began a conversation around, OK, AI is very powerful, it can do lots of great things, but what are the implications of AI and things like the kind of data that you're using to train it? Can it perpetuate biases? And so that's when the conversation around regulation really began to take root. And then a lot of conversations about, you know, how – and still today – how do we address the issues that underlie the data that's fed into these AI systems so that we can make sure that it's representative of the use cases, that it's not perpetuating harmful biases, and so forth?

And then that led into a number of – a recognition of a number of other challenges around AI, which is that, well, how do we even understand what's happening? It's a little bit kind of black box. And so, well, you say, well, let's be transparent about using it or let's have it explain what it's doing, and that way we can more easily pick up on some of these potential harms. But then we recognized it's not so easy to do that from a technology perspective. And so DARPA, for instance, began investing in next-generation AI where you can have more transparent and explainable AI systems, and that still is a challenge for today.

So then you just kind of fast-forward three years or so, and then that's when the generative AI tools sort of hit the – hit the market, hit the internet. And now, because it's – they're freely available – tools like ChatGPT and Gemini and others – that has now very much opened up the possibility of using these kinds of technologies in ways that previously had not even been imagined. And it's really ratcheted up, again, the power of these tools to do things that humans are often not very good at.

And so now you, again, have this – the power – the potential for innovation in so many different ways, like, for instance, democratizing creativity in the sense that, you know, you may not be very good at creating that marketing brochure that your boss told you, you need to create by tomorrow, but ChatGPT or Gemini or one of these tools can give you a good starting point, and then you can work from that to move forward. So lots of applications like that of how these tools can be useful.

At the same time, now we're seeing yet another kind of concern, which is, well, how are these systems being trained and what is the material that is going into these systems? And so that kind of brings us to the topics of today, which is these large language models, these generative AI systems, are being trained on information that's available by and large just on the internet. Effectively, it's reading the internet and using that to train.

And so then that leads to the challenge of, well, how do we ensure that the owner of that data that's being used in these systems is being appropriately recognized, compensated, and they have consented for that? And that has

now raised a lot of the copyright and patent challenges around what is the right way to think about this.

And so I think that's certainly a lot of what our – what our conversation will be about. But I think in terms of the trajectory of AI and why it's exploded onto the scene now, it's just very much because as these tools became more and more useful and you could do amazing things with them, that sort of practice with using the technologies uncovered a number of these potential harms – things like using copyrighted material in a way that people aren't sure if it's fair use or not, going back to the potential of the data itself perpetuating biases.

Another concern is around privacy. You know, already, absent generative AI, our data is already being collected, say, as you click through different pages on the internet. And that click trajectory might then tailor advertisements to you, and maybe you don't like the privacy implications of that. So there are a lot of these kinds of issues that arise I think largely from the data that's being used to train these systems.

So that's kind of the lay of the land of where we are and how it came to be that these are the issues that we're grappling with.

Dr. Gupta: Yeah. And the big confluence of those three factors – compute, big data, and the transformative technology – coming together and just unleashing.

Dr. Parker: Yes. Yes.

Dr. Gupta: And by the way, panelists, jump in with any additional points. But I was just going to move from the two points that you noted that are related to artificial intelligence is, one, creativity, so, you know, what does that – what does the new system of inventing and innovation look like with AI-assisted inventions; and the other one is of data, who's – what are these data inputs, whose data is it, what do we do about that. So maybe we start by unpacking that.

So I'll move to – first to Rama and then to Phil just to talk a little bit about the question of AI-assisted innovation. So, Rama, I'll start with you. Patents, right. A lot happening in the world of patents. What does that look like?

Rama Elluru: Thank you, Kirti, and thank you, CSIS and LeadershIP for including me on this perfect panel.

As Dr. Parker noted, AI capabilities are growing exponentially by the day, at least by the week, and so these capabilities are providing better insights into a broader array of subject matter. They are providing new ways to use the AI.

So, for example, AI is being used in drug discoveries. So, for example, Novartis' JAEGER system was trained on over 21,000 proprietary molecules. It was seeded with three molecules and asked to provide new novel anti-malarials and it generated over 200 of these compound and they did work on it to see which ones had the characteristics they wanted, low cytotoxicity, and they did more experimentation, of course, and then went on with that. And Corey, if you're watching, I hope I got that right.

But also it's being used in antibody identification so you can tell the AI this is the target that I want an antibody for and so it's been trained on what that target looks like, what its, you know, dimensions are – what its 3D dimensions are – and what kind of antibodies can bind to these targets, and it is doing this much better, faster, than humans can because it is trained on more data, it understands a lot more than a human can or what a limited number of people in a lab can.

And another example in the tech industry is chip design. So engineers have to design chip floors to see how the components are going to be placed to maximize space, power, wire length, et cetera, and AI is being leveraged to – for chip design. So and maybe not generative AI but it's still the same AI architecture that enables generative AI large – like large language models.

So what does this mean for patents? I think there's several issues that AI raises and challenges for intellectual property rights. One is subject matter eligibility as – and I know it's been talked about a lot but we know we have subject matter eligibility – I would call it a mess – here in the United States and it doesn't help for AI because software already is vulnerable to patent protection and AI at a very high level can be thought of as an abstract idea. So how are we going to protect AI as an invention of itself?

But getting back to AI capabilities, a lot of people are using AI in the invention creation process, as I said. So when that happens who is the inventor? So is it – you know, our laws say that we can only have natural persons as inventors so – but what does that mean? Who in that process when I'm using AI, for example, in a wet lab to, you know, give me suggestions on drugs or molecules that I should be exploring?

So the USPTO did issue guidance on patent inventorship and they said that AI-generated inventions are not categorically unpatentable. They did say that natural persons have to make a significant contribution to the invention to be able to qualify as an inventor, and then they provided, you know, principles and we can go into that later and this, you know, is a great, great start.

I do think that this means that there is other patent challenges. There's now a new patent challenge about did a human make significant contribution when AI is used in an invention creation process and I think this issue will come up during the patent examination process itself.

But I think it'll also come up in litigation. I think this is a new hurdle to patentability and enforcement. I think obviousness is also an issue because when the use of AI in creating inventions becomes more commonplace. And you just said that it was obvious to use an AI to help you come up with this answer.

I think enablement is an issue when AI is used in the invention creation process because, as Dr. Parker mentioned, we don't really know very well how these systems are working so how can you explain that you own – you know, you had written description or you owned this – had this invention in mind?

And then I think that AI can also generate prior art. You can push a button, give it ideas, say generate prior art, put it into the public domain as a defensive mechanism. So I think there are a lot of issues too that we're going to have to deal with. I call it job security for the IP ecosystem.

But I think there's also great measures. I think the USPTO guidance did a great job with a guidance and, you know, they have said it's going to be an iterative process. So I think that we need to make sure that our IP system is sustainable, that it is – you know, it is protected so that – and that these laws don't – these guidance and these rules around patent inventorship and all these other issues that I just talked about don't undermine our patent regime.

Dr. Gupta: OK. Can you give us some – like, are there examples out there where these issues have been stress-tested, where somebody applied for an AI assisted patent and how did the Patent Office react and how did the courts react?

Ms. Elluru: So in the guidance the Patent and Trademark Office gave two hypothetical examples. One was drug discovery and when – you know, for example, if a person just uses an AI model to ask for suggestions on which molecule should I start with, who is the inventor in that process? Is it – you know, can it be the person who built the AI model? Can it be the person who prompted the AI model? Is it the person who got the output and did more work on the output to see if it was – what was intended and will work for – work as intended.

Another example they gave was in the tech sector of finding I believe it was an axle for a car and finding new material to use. So those are hypothetical examples but I think we will see more real-world examples as time goes by.

Dr. Gupta:	Yeah. And let me, you know, quickly turn to you, Phil. Same question, right – AI-assisted invention but creation in the copyright space. What are we seeing just on that question before we go to data and dedication? What are we seeing and how is that being handled right now?
Phil Hill:	Well, just like the Patent Office the Copyright Office also issues guidance, and I'll talk in a little bit about the path by which we got to that guidance.
	But, fundamentally, the Copyright Office has given some examples of things that would need to be disclosed and disclaimed because they are the unaltered products of generative AI and not the products of human authorship.
	So, for example, using generative AI for ideation purposes or for a functioning thesaurus or to optimize for a particular audience, the Copyright Office has said that falls into the bucket of in the copyright space call this assistive AI. So this is something that has not really generating an expressive product that a human being otherwise would, but it's helping the human being in some way.
	And then for the generative aspects, you know, the market is changing daily but we see text generation, image generation, code completion, multimodal systems coming out. So we see those all the time and as those produce expressive works that a human artist would normally have to do those are some issues that we end up having to deal with in terms of teasing apart the creative process and finding the human authorship in the process of creation.
Dr. Gupta:	And maybe we ought to talk about this specific case that people might have heard in the news of a photographer taking a picture and translating it into a painting that was inspired by Van Gogh and the Copyright Office denied application of –
Mr. Hill:	Yeah. That's one of the later decisions that the Copyright Office had. And interestingly, that same applicant had applied for registration for a similar work or the same work in India and had gotten a registration preliminarily – this was "Suryast" – and then it was later cancelled in India.
	So in the United States this was a particular model that was a bespoke model – I think it was an academic researcher – where you take an image from some source – in this case it was "Starry Night" by Van Gogh. They took their own image and then you can set the parameters for how much of each of the two contributions should be manifest in the output and there the Copyright Office said, OK, there's not enough creativity here that we can source to the human operator and so then, therefore, the application was denied.

Dr. Gupta:	So, Lynne, going back to you. So between the spectrum of, you know, taking a picture and modifying it to R&D inventions in biotech and semiconductor chips there's a spectrum and the question that's asked is a broad one. When there is an invention that's AI assisted where in this spectrum are you going to draw the line of what's patentable or copyrightable and what's not, and it's a difficult question.
	And let me just, you know, sort of frame this for our conversation a little bit. It's not a new one because we saw the same question in software in the past and we know – actually, I would like to ask you this from a very technical perspective. How do we know – how do we define what's software and what's AI and what's patentable and what's not?
	Doesn't this create a really tricky set of questions?
Dr. Parker:	Well, AI, is software. It's a special type of software. I think, you know, if you get a roomful of people who are AI experts in a room and say what is AI, you know, they're all going to have a different opinion.
Dr. Gupta:	Right.
Dr. Parker:	Actually there is a legislative definition of AI and so I think that's what new regulations and so forth are defaulting to. But it's not particularly an operational definition.
Dr. Gupta:	And it's a moving target.
Dr. Parker:	It's a moving target. And so I think there is certainly a challenge of saying this is AI. You know, when I was in the federal government and doing things like trying to scope out – just as an aside – things like how much is the United States spending on AI research and development. We also did it in a scoping statement of saying anything within the scope it's up to the agency to decide what actually is AI.
	So I am not optimistic that we'll be able to have any sort of regulation or policy that definitively says this is in scope, that's out of scope. I think it's going to have to be – it's one of these cases you recognize it when you see it kind of thing.
	But, you know, I'm not really an expert on, you know, what would be patentable as it relates to the spectrum of software versus AI. I think when it comes to the human contribution that is – you know, at one end if it's purely some AI out of the sky suddenly generates something I think there's consistency that – consistent opinion that that is not patentable, obviously, because there's no human involved there.

But where you draw – how much human involvement is – whether it has to do with software or AI I think that's going to continue to be a challenge that I suppose is going to have to be determined on a case by case basis.

Dr. Gupta: Are you concerned about this, Rama? Because, you know, when we saw this for software patents, for example, there were some big cases in the Supreme Court that led to this question of patent eligible subject matter, what's patentable, what's not in a software patent, and it was a tricky battle because, you know, I've done a few empirical studies in the past and I've seen that, first of all, what you define as any category of patents – software patents yesterday, AI-assisted patterns today – is a changing definition and it can be a really broad one and it can implicate a lot of technologies that are critical for national security and making them often not patentable.

Is that a concern you're facing?

Ms. Elluru: I think that is a concern, right? I mean, just like that first panel said innovation is really important in this global competition. IP has historically been a lever in fueling that innovation. So if we want to be – if the United States wants to be leaders in artificial intelligence and related technologies we also need to be leading in our intellectual property rights to make sure that we are protecting them, to make sure that is still being used as a lever for innovation.

I looked at other words for incentive during that first panel but so it is that enticement. It is the motivation. It is what allows you to make those investments which are sometimes incredibly burdensome and so it – again, as that first panel said it is for – it is especially important for the small or medium sized players who have hurdles getting into the market, getting into the ecosystem.

I agree that trade secrets is being used as an alternative to the patent system but so is open source. I think open source is great for many reasons – democratizing, access, et cetera – and we need a balance of intellectual property rights.

It can't be just trade secrets and open source. There needs to be a place for a strong patent system for those who need it, for – you know, because the trade secret option as well as open source is not going to be an option for everybody. So we need this option for some people who really need the patent system as the enticement to make the necessary investments in resources and money, especially for those technologies that are much longer term.

	And so it is a really upfront heavy risk. All you have at that point are intellectual property right protections. So I do think that is – it is very important and it's something that we really need to make sure we get right if we are going to lead in innovation.
Dr. Gupta:	Anything you want to add, Phil?
Mr. Hill:	The same concerns apply for copyright and I'll say that in general the Patent Office and patent law has some built-in safeguards as well because they're technical. There's a long process. You need expert examiners and attorneys to represent you for the patent system.
	For copyright there's an accessibility concern in that anybody can own a copyright. A child using an iPhone that has ChatGPT on it can say, give me a software program that will say "hello, world," a very simple closed universe of code that they can have. You know, is that copyrightable? Is that not?
	Those narrow use cases might be some of the use cases that the Copyright Office sees more often, often in terms of, you know, generating images but all the way up to huge code bases with millions of lines of code and armies of engineers and use developers all over the world, these types of use cases and the issues that go along with the protectability of copyrightable works and then what you need to subtract out, which we'll get into in the guidance in a moment, raises great concerns in terms of protectability and the ability to invest and leverage IP that you create.
Dr. Gupta:	So speaking of the word incentives I think that's going to be today's word. (Laughter.) Where is Patrick?
	But speaking of incentives, you know, one of the issues that now we are dealing with in the copyright or this space is basically the use of content data and litigation in the copyright space as the use of the gen AI models has taken off, which uses millions of data inputs, right.
	So there's a ton of litigation about whose content – is it potential copyright infringement and the value of the content. So, Phil, I would love to, you know, walk us down the litigation landscape today.
	What are the key questions that are being asked right now?
Mr. Hill:	Yeah. Well, I'll say from a litigation standpoint, you know, my team and I we're tracking about 60 litigations that deal with AI or generative AI around the world. Most of that's happening in the United States and most of that's happening in copyright so I'll focus mostly on copyright.
Dr. Gupta:	Sixty? Six-zero?

Mr. Hill:	Sixty. Sixty. Six-zero, that's right. Depending on how you count. So but we've seen – you know, since the first cases were filed in the copyright space, you know, a steady increase in the sophistication and focus of the claims, and so one of the big questions that we're seeing as a pattern is the parties and the judges are increasingly getting under the hood and disaggregating the process, looking at how things are being used, by whom, when in the process, and by role and by stage.

So one of the key decisions came out as a motion to dismiss decision in Andersen v. Stability. That's a putative class action bought by an artist against Stability for an image generation AI, and one of the important things for that case when they went to court there wasn't a clear distinction between which of the defendants did what.

So one of the defendants was Stability, which was alleged to have been responsible for a company going out and assembling 5 billion images to train on that's – (inaudible) – and then also they developed their own foundation model, and then also deployed their own model. And then, as co-defendants in the case there was DeviantArt, which in-licensed Stability's foundation model and then deployed its own model with Stability powering it.

And then there was another entity, Midjourney, which was at least a deployer of something that relied on Stability and then the judge said, I can't tell whether you're alleging that Midjourney also did some supplemental collection and some supplemental training on top of that, like fine tuning is the term that we would use in the AI space there.

And so the judge said, OK, I'm going to dismiss this. You have leave to amend and one of the things that you need to do is more clearly articulate who you're saying did what in the process. And so that's one of the fundamental questions that we end up seeing a lot here is who did what and how do these things work under the hood?

Another issue that came up with that is right now anyway or at least based on the time frame at issue the way that deployers and end users of the generative AI models like Stability would have to work is they would have to submit a prompt and then there would be access to a hosted version of the foundation model and so one of the things that the judge said is, well, how does that implicate copyright because copyright requires a copy, requires a distribution.

And so if you're only providing access to something that exists on a server how can you say that at least the deployer defendants have done something that implicates a copyright right. And so there's leave to amend to address those issues as well.

One of the big questions that has come up and that we do have some determinations on is a question of whether the – we call it the causal connection theory of copyright infringement as viable and every court to address this so far has said no so – and this is a theory that enabled a lot of the initial complaints in the copyright space or at least they were based on.

So in the Andersen case, again, those plaintiffs said, we don't need to allege that the copyright that I own is substantially similar to the model or the outputs of the model because they would rely on this idea that because there's an underlying instance of copyright infringement that was used in the collection and training of the model – that's sort of a poison fruit analogy – everything that comes afterward that touches on that is necessarily an infringement.

So they didn't even include a side by side in the first instance, and the judges have all said, who have addressed this issue anyway, have said, no, no, in order for there to be copyright infringement you must show that there is a copy or substantial similarity.

So the motion to dismiss was granted. They had the ability to come back, amend their complaint, and say, OK, here's some substantial similarity allegations and they have supplemented that to a degree.

I think those are the some of the major issues that we're seeing now. Of course, we're seeing some previews of the fair use defense. We may get some early decisions on that as well particularly because there's an opposition to a preliminary injunction motion in one of the cases so we might get an early indication with that.

Otherwise, we're seeing some of the defendants fast forward to a fair use defense, mostly as atmospherics in some of their briefings of the motions to dismiss and other preliminary procedural mechanisms. But they're not the basis for the determination of the decision.

Dr. Gupta: Before we get to fair use, I just want to unpack this a little bit on the – with the technology input here. So what you're basically saying is I think it's, you know, for sort of generally audiences now tracking copyright litigation, which I am too, by the way, but this cartoon comes to mind where you have these – you know, there's artists and writers and painters and musicians that are fishing and each one has a fishing rod and is taking a fish and whatever they catch they're collecting it in a basket behind them together.

And then there's another player who's wearing a jacket called GenAI who is putting the fishing rod in the basket with all the collected fish. So I think that basically draws the picture of what's happening and then the question is

there's a lot of this content that is being used to train these models. There's a lot of input.

So how do you determine? I think, you know, Phil has laid out how this is drawing out in litigation and the question that is being now synthesized is, well, what specific training set did you use and how did that specifically relate to your output.

But can we do that from a technical perspective? How hard or how easy is it to do that?

Dr. Parker: So I think at a theoretical perspective what you would like to do is say, here's a piece of input and here's the way it weaves itself through this AI model and here's where it kind of influenced an output and, you know, theoretically you could imagine maybe ways of using provenance or where data has come from and tagging it in certain ways so that you can determine that.

But you're looking at trillions of inputs and there's not a direct – you know, there's no copy of it within the model that's being built. What these generative AI models are doing are looking statistically at these trillions of inputs and are coming up in ways that we don't fully understand with incredibly complex models that are predictive of something that seems realistic. But nowhere in there can you find that input just, you know, kind of pasted. It's not doing a collage of things, for instance.

And so I think the challenge is can you do that kind of provenance checking through an AI model and do it in an efficient way. Not every input to the system actually influenced how the model ended up being formed and then it may not have been used in terms of any particular output.

So I think the challenge is always you have inputs. They may or may not have ever influenced any sort of output, and being able to trace that through the system. Technically, you could sort of theoretically imagine how you would do it. I think from a practical perspective it's not going to be easy. It's not easy to do that.

Dr. Gupta: Yeah, Rama, you were saying something?

Ms. Elluru: Actually, I have a question. So if we can't technically have provenance to sort of connect the dots between how the input got into the model and to the output are there other options – policy options that we should be thinking about?

Dr. Parker: Well, I think certainly transparency is always a good thing and so in theory if you want to say this is what was the input to the system that's just kind of a

starting point. But I don't think it's a very satisfactory answer because you, therefore, as I just said don't know how it was used.

I think certain companies are exploring only using licensed data. I think Adobe is one that only uses licensed data. I think the challenge you run into that is that by definition you're going to be restricted to the data that you use and now we have this flip – kind of this duel between you want to reduce bias and the way you often reduce bias is by having more input.

Bias typically happens because you don't have enough data about sort of small or sort of outlier cases or what often – and the technical terms it's called small class versus large class, meaning the majority of your data is about one thing but you have some data that's very important but you don't have as much of it.

And so the – what you're trading off, I think, if you only use approved licensed data is that now you're more likely to end up with some sort of biased results because you don't have the broad spectrum of input. So I think achieving both policy goals at the same time is not immediately straightforward.

Mr. Hill: In the copyright litigations at least there's allegations about how you can derive some of the training materials from either the model or the outputs and so one research paper by a Google engineer named Carlini has been cited over and over again in the complaints and basically that research paper says for image generation anyway that his team of engineers could go into an image generation model and using certain techniques, so not using it for the outputs but just go to the model, derive some of the original source material that they have, about a hundred or 200 instances of images that they say they could get from back forming.

And so one of the things that the Andersen plaintiffs did when they had their motion – after the motion to dismiss to amend they amended their complaint to add additional citations to this article plus other research that says, OK, there is some possibility that the model itself has memorization. Sort of the counterpoint to that this is just weights and biases, which is what some of the developers might say, would be on the other side that this is a sophisticated and probabilistic compression algorithm that you might say 99 percent of the time this word is followed by that word or 100 percent of the time this word is followed by that word.

Relatedly then the second way that people have been trying to divine whether certain things were used was by consulting the outputs. So in the Concord v. Anthropic case these are music companies that have sued Anthropic over lyrics and they said, OK, give me the lyrics to a "Fresh Prince of Bel-Air" and then on output the entire set of lyrics for "The Fresh Prince of

Bel-Air" came out. And then they would say, OK, give me the lyrics to a song about somebody moving from west Philadelphia to California, in the '90s and they gave the lyrics to "Fresh Prince of Bel-Air," more or less.

And so if it was really just probabilistic and it was really just weights and biases then you might get a bunch of lyrics that surprise and delight you in various ways but are not exactly 100 percent the lyrics to "Fresh Prince of Bel-Air," so the story goes.

And then the last way for – especially for smaller models like you had mentioned there are ways of attaching provenance but mainly that's for where you've implemented the generative AI technology in a particular way in the deployed use case. So you would have a foundation model which basically is understanding language or whatever the modality is that you want.

Then you might have a set of inputs and those inputs might exist in a freestanding database like it's consulted or it may be the user adds inputs into the prompt material and those things that might be ways that people are looking at to say, OK, this is what I'm going to use as part of the output – the generative AI will say that – and then include a citation to a particular set of things that were used and incorporated into the output, things like reports and things like that.

Ms. Parker: Yeah. I think one of the challenges sometimes is that these models are just – you know, they're sort of reading the internet, right, and so they're getting information that's off of a page at a URL at a certain point in time.

But, you know, what's on that page the next day may be different and so keeping track of how all of these things are, you know, at the moment that you accessed it what did the page look like effectively could mean that you have to reproduce the internet at a moment in time all over again. And so the technical viability of that is pretty low.

Dr. Gupta: Yeah. And because you're seeing, you know, trillions of inputs, you know, probabilistic models and there is this question about fair use are you – if input is being used for training these models is this copyright infringement or not? Is there a question of valuation that occurs?

And I think, you know, Phil, I would like to ask you a little bit about the music industry. You've been involved in the music industry, licensing litigation for a long time. Eventually they came up with business models, licensing models, to aggregate content and to be able to license it.

Is that a sensible possibility? Is that happening? And if it does, the analogy only goes so far because valuation – how do you place a value on different

	kinds of content when there is so much input that is being fed to train these models? What's your read on that?
Mr. Hill:	Yeah. There's a really interesting evolution of the discourse here. So I'm going to the older examples you're talking about here. In the Copyright Office's process of listening sessions they had four sets of listening sessions with stakeholders who came in, including from some of the major record labels and publishers as well as other industries, and those representatives basically said this isn't like piracy – online piracy. What we want to do here is find a way forward. We're not trying to shut this down. We want to find the right way for credit and compensation.

And so they're finding a path forward and that's a useful data point. During those conversations the rights owner said, we want to license these things. We want a free market solution to this. And they reported that when they had those discussions the answer on the developer side was this is fair use. So you can get a license but that license value is zero because it's not infringing.

Today, there are content – there is content being licensed for the development of AI. The question really is how useful is that and for what use cases because in some instances you need to train on lots of things – a huge amount of things. In other cases you need less depending on what your objectives are.

My team, for example, we are doing content licensing for AI uses in training and as inputs and supplemental inputs in various other use cases as well for music and print and everything else under the sun. |
Dr. Gupta:	So that's happening?
Mr. Hill:	It is happening, yeah.
Dr. Gupta:	Rama, you were saying something?
Ms. Elluru:	No. I did want to comment on data if that's OK.
Dr. Gupta:	Yeah, that's – I was exactly getting to that. You know, data is a critical input to AI. What are other countries – like, what are the big issues there in what other countries are doing?
Ms. Elluru:	So aside from copyrightable data, you know, there's data that's really critical to AI, especially in the health care space, drug discovery space, curating really big quality datasets that are so important and critical to some of these models or even creating synthetic data when the real data isn't going to work for whatever purposes – bias, privacy issues, et cetera.

What are the motivations – sorry, incentives to create those types of datasets? We don't have data property rights other than the copyrightable data here in the United States. China – so back in 2021 when the National Security Commission on AI issued its final report it included recommendations to strengthen the United States IP system especially when China is making IP a priority in its national strategies and strengthening its policies and – IP policies and strategies.

And one of the recommendations there was that we should explore data intellectual property rights, what that would look like, why we need them, and so on. And that was in spring of 2021. By the end of 2021 there was reports that China was exploring pilot projects for data intellectual property rights across its innovation hubs and this year I've seen news that the – China's intellectual property office has taken up the effort to see if data intellectual property rights would create necessary markets for data.

So I think that is something that we should explore here especially, perhaps, in certain sectors like health care diagnosis, drug discovery, to see how we can democratize access to data so that big quality datasets, synthetic data, et cetera, is not just sitting in a few companies.

How can we get that data? How can we create those protections and motivations to democratize that data to other smaller players who may be able to help fuel the innovation ecosystem?

So I think that is something – yeah.

Dr. Gupta: That's a really good point, yeah. And you know, speaking about what other countries are doing, I just want to go back to this broader question about – I think Carol Goldman raised the point about this cannot be – none of this can be done unilaterally. We need international cooperation, and in terms of, you know, taking a step back from the IP angle of AI, you know, writ large, like, a lot of countries are looking at artificial intelligence. They're looking at the implications for policy and regulation.

I have this data point from Jim Lewis who sits here at CSIS. He wrote that there are 37 different AI-related laws passed in a hundred and twenty-seven different countries in 2023 alone. They're asking and answering a lot of questions, granted, not just related to IP, right, on safety and security and bias and so on.

But it really raises the question about this international patchwork of laws and regulation and how to navigate that. Is there a role for the United States here? What are we doing? What can we do? I would like to ask that question to all of you.

Dr. Parker: I'm happy to start. I think, as the previous panel mentioned, maybe thinking about it in terms of regulatory regimes that harmonize with each other as opposed to a single global regulatory regime I think is kind of a realistic goal.

You know, going back to 2019 when the OECD countries, certainly led by the United States, agreed to a set of principles about the appropriate stewardship of trustworthy AI I think those principles – now they're five years old but the overarching perspective on how to use AI in a responsible way I think most of the world agrees and certainly in the United States I think there is a strong bipartisan, bicameral perspective on what those policies ought to be, what the characteristics are of AI, and the trustworthy use of AI.

I think when it comes to taking those and turning them into regulations it's going to be different, I think, for different regulatory regimes just by – almost by definition. I think one super important advantage that the United States has in that regard is that we already have lots of agencies that have regulatory authorities over particular use cases like FDA and health care or DOT in the transportation space, and so forth, and those are the agencies that can dig in and think about what are the use cases that are high risk.

And this is then harmonious with what the EU is doing with a risk-based approach, meaning you don't have the same regulatory oversight of use cases that are low risk in terms of impacts on people or safety and security, national security, versus those that have a high risk or even some use cases that you might want to prohibit altogether.

And so I think, certainly, with the most recent executive order from back in October I think there's progress in pushing those agencies that have regulatory authorities to actually come out with a guidance so that we can have sort of a U.S.-based approach to defining what that regulation should look like in AI that's consistent with our regulatory setup, which is not the same as what the EU has.

But I think the overarching principles of safety and security and fairness and so forth are the same. So I'm somewhat optimistic that we can come up with harmonious regulations eventually.

Dr. Gupta: Rama, you want to –

Ms. Elluru: Sure. I completely agree. I mean, I would say the U.S. is leading right now but not by much in AI and other technologies. But we need to make sure that we sustained that lead and we can't do that without our partners and allies, and to be able to have data flows, trade agreements, et cetera, we do need harmony in our regulations.

I did want to make one point to Dr. Parker's recognition that the EU is using a risk-based approach. The United States doesn't have a risk-based approach and what would one look like. So the Special Competitive Studies Project did issue a framework for identifying highly consequential AI use case whether that's beneficial or harmful – what does that look like.

So the EU AI Act only concentrates on risks but here in the United States we want to harness the benefits of these technologies. That's why we are interested in innovation. So how do we get that balance between regulation and innovation? And one way to do that is to really focus our regulatory efforts on those AI use cases that's going to have the most significant impacts on our society, again, whether that's beneficial or harmful.

So what does that look like? You know, we had – the framework that SCSP issued lays out the different categories of benefits and harms and identifies specific harms and benefits in each and has a way of assessing sort of a heat map of, you know, what kind of impact is this going to have on society.

So I do think the United States also needs a risk-based approach or some sort of highly consequential-based approach and so I think that's where some of our efforts should focus as well.

Dr. Gupta: Yeah. Phil, do you have something to say about this?

Mr. Hill: Yeah. On the international comparative front to say that the U.S. Copyright Office has been taking comments and having listening sessions and having webinars about international aspects of IP for the purpose of considering harmonization.

I'll say that other countries have had final determinations, or at least preliminarily at the lower court level pending appeals. So there was a decision in the Czech Republic last week dealing with ownership and infringement claims for AI-generated images.

China has at least a handful of decisions dealing with ownership of generative AI outputs and also one infringement case as well. And so the United States, you know, we are in the process of litigating some of these issues but it's going to be a long time until the litigations finish.

There's policies that are happening now that people are trying to organize around. But to the extent that it takes a while to get to our final determinations on what the copyright regime should be in the United States other countries are also setting a bar that if we harmonize to may answer the question for us before we get there.

Dr. Gupta:	So, you know, as you guys are saying this, I think these are – the technology, it's such a large-scale – it's changing everything. So it really has an impact on the – word of the day – incentives and how we are going to innovate and create into the future.

So any of these IP policies around AI will have sort of these long-term and rippling effects through society. That's why they matter in this gray scale area.

With that, let me just open the floor to questions from – and comments, participation from our room here and we'll get back to the panelists. It's a new technology. Any question is a good question. (Laughter.) We have a couple here. (Laughs.) |
Audience Member:	Well, I do not have a question, but I just have a comment, because at the beginning it was said that it's difficult to define AI because things are changing. But I think that we can talk about two definitions for the general public. I would say – if somebody asks me what is AI, I would say in general it is the attempt to translate some elements of human intelligence into computer tech. That's it. Now, if somebody said, no, no, technical, OK, it's math plus data – (off mic) – algorithm plus training to get a software; plus a final decision-making, that's AI. The grouping of math/data training to get something that you as the creator want AI to do. That's it. So simple. (Laughter.)
Dr. Parker:	So let me comment on that. So –
Audience Member:	Could you repeat the question?
Dr. Parker:	Oh, yes. Not everyone heard the question.
Audience Member:	It was – it was not a question. It was a comment.
Audience Member:	Or a comment.
Dr. Parker:	It was a comment, yes. The comment – she was suggesting that maybe there is a definition of AI that's math and data and algorithms –
Audience Member:	Training –

Dr. Parker:	– training, and then – and it's doing something that is – traditionally you'd think would require human intelligence. And I think most people would sort of loosely agree with that.
	But let me give kind of a bit of a counter example to focus our attention on what the goal is – what's the goal of defining AI – and I think in many of the regulatory conversations the goal is more about making sure that people aren't harmed by the technology in some cases.
	And so there I had the opportunity to testify for the – before the Senate Homeland – HSGAC – Homeland Security and Government Affairs Committee a few months ago and another person on the – another witness was a fellow that used to be ACLU representative in Idaho and he had been litigating for a very long time harms that were caused to an individual whose – some social benefits that they were planning – had been getting had been reduced by a substantial amount and they were told, your benefits are going down, and they wanted to know why.
	And they asked the office – you know, the government office, why are my benefits going down and they said, well, just because. I'm paraphrasing it. You know, and they kept going back and, well, why are they going down, and eventually it got to the point where it was a formula in an Excel spreadsheet that nobody understood and so everyone in the office was saying this formula says so. I'm paraphrasing again.
	But the gist is it was an Excel formula. The vast majority of us would say that's not AI but it was harming – it was being used to make decisions about people's access to resources that was harming them.
	And so I think when you're thinking about regulatory challenges and what it is you want to regulate about it's any kind of sort of automated decision-making that is potentially impacting people in a negative way, whether it's an Excel formula or something that is more like a neural net. And so I think that's why for academic purposes, you know, you might want to come up with a definition and certainly in legislation there is a definition.
	But at the end of the day, if the goal is to minimize the impacts of technology on people as it relates to things like access to resources we have to be sure that we're not scoping out things that have exactly the same impact even though we may not call them AI.
	So it's kind of a complicated territory when you couple it with the regulatory goals.
Dr. Gupta:	Very interesting.

Audience Member:	Thank you for a fascinating, very interesting discussion. But you have been using copyright and patents sort of flowing back and forth. But the two regimes are very different and all the economics of it, of innovation, tell you that they have very different ways of affecting the incentives, since you've mentioned.

So could you unpack that? But there are three parts. The second part to that is even in copyright we've been talking about data and other parts of copyright, let's say – of coding, let's say. Well, can you copyright codes? Because 90 percent of people use Copilot, or 99 percent of software developers, people working on AI, are using that. That becomes another issue, because copywriting data is much more in the news because of the news media, but from an innovation point of view perhaps the problem with Copilot will be a bigger one in terms of code.

Lastly with that, in terms of copyright, you know, the Andy Warhol case – the use case versus the transformative – that raises very fundamental questions, and that changes the valuation of different parts.

So I think it would be useful to unpack these in terms of understanding what the economic impact is, because I think the reason the Chinese have gone further ahead in terms of getting their arms around this complex subject is they have not let it be tied into the – some legal technicalities rather than taking the bigger picture and then seeing where are the bottlenecks. Don't you think we need to do that ourselves? |
| Ms. Elluru: | So I think the biggest inconsistencies – and would love your thoughts on this – between current patent law and copyright law is just the, you know, patentability of AI-generated inventions. I mean, it's not categorically unpatentable. Whereas I wonder if the copyright approach is a little inconsistent with that. And maybe there is – if it is there's good reason for it given the economic and policy reasons we cite between the different – between the Patent Act and the Copyright Act.

But I do think that's something we should address because we don't want any inconsistencies to undermine the IP regime as a whole. |
| Mr. Hill: | Fundamentally in the copyright space we have questions of ownership and then we have questions of the outputs and so there's a lot to unpack in your great question there. But one of the big fundamental differences between patent and copyright is that a copyright infringement is implicated from the moment it's used – it happens, the moment a copy is made, the moment the distribution is made, the moment a public performance is made, whereas an infringement for patent you have to practice the patent somehow without authorization. |

And so that gets into a different regime because maybe an AI produces the description of what the patent is but somebody still has to put that into practice, right, whereas a copyright infringement may occur just from the moment that something is copied into a training set or is latent within a model or is produced on output.

It doesn't require the additional step of putting it into practice in the industry or the marketplace, similar to trademark. Trademark has to be used in the marketplace to identify a good or service and so you have this air gap between the model and the ultimate use case.

Dr. Gupta: Yeah?

Audience   Either a question or a comment but I would like to brainstorm just a bit.
Member:
One of the general themes that I've noticed in this panel discussion is the difficulty of distinguishing the human inspiration from the artificial intelligence inspiration and I feel like much of this could be addressed if we incentivize both a centralized artificial intelligence model and a local artificial intelligence insistence and what I mean by local is unique to each individual user's device, a(n) independently encapsulated artificial intelligence assistance.

And what this centralized and localized distinction will be able to do is form a relationship with human interaction, an actor. So the local AI can serve as an activist for the humans that it represents and can negotiate with the centralized artificial intelligence model, and having established that personalized relationship at the local level then we can address if the inspiration actually came from that individual person versus, you know, was this inspired from some centralized collection of inputs.

Just brainstorming an idea there. Thank you.

Dr. Gupta: Using AI to solve the question of identifying who the inventor was.

Ms. Elluru: Yeah, I think that's a difficult one because I'm not sure size matters in determining how much the AI has contributed versus the human has contributed in that spectrum. I do think that's where the future is headed where we're all going to have some sort of AI-enabled assistant.

I think that maybe your point might be that it is easier to sort of define who – what the human contribution is in those localized AIs as opposed to sort of centralized AI. I think philosophically we really need to be empowering the humans to make sure that the humans are controlling the AI as well as those – you can't incentivize AI. You can only incentivize human. So we need to

make sure that any sort of intellectual property rights are inured towards humans.

But, yeah, I would have to give that some more thought.

Mr. Hill: I think in general for all of these issues the devil is in the details and you get taxed on what you don't know, and so that makes it a little bit difficult as well.

I guess another, like, bumper sticker that you might throw out here is AI, is it both the problem and solution to all of the problems that we have now? So maybe that's one way of doing it.

But going to the question of whether you can have a local instance sort of mediate the interaction with a larger instance. Some touch points for that might be in the U.S. Copyright Office they rejected an application where somebody had said I myself, a human being, use more than 600 successive prompts to an image generation AI so isn't that enough human creation, those 640 some-odd prompts, and the Copyright Office said no, it's not. It rejected the application.

And then the Czech Republic decision that came out last week, as I mentioned, that prompt was something along the lines of give me an image of people signing a contract in a board room showing only their hands, and the court there said that prompt in and of itself isn't enough.

So you go in, you see what the prompt is, and if you have the records of the prompt then you can see, well, does this have enough expression in it or not. You'd have to have those and then you'd also have to be able to tease apart especially when things become much more complicated and then maybe you could have AI mediating the prompts as you put in sort of a human-generated prompt and then it interprets it into better instructions that feed the prompt as well.

Dr. Parker: That's coming, too.

Mr. Hill: Yeah.

Dr. Parker: Yeah. Just one observation. I think, you know, the last five years of AI development have shown us that we have to think dual use or triple use and so if you – that would require, I think, some research and development to create that kind of capability.

And then I think, well, what else could it be used for, and I think it could also be used for an employer to see how much work your employee is actually doing. So, you know, you get into these gray zones of, you know, do you

	really want to go down that road of having an AI system that kind of tattles – (laughter) – in some sense on what the human is actually doing in sort of a dual use kind of perspective.
	So I think we just have to kind of think through the ways that these kinds of developments can be used in a harmful way as well.
Dr. Gupta:	So you have the unfortunate task of being the panel between everybody and lunch. So I'll maybe finish by the – what Walt said, a quick fire drill with a quick takeaway: key questions we are facing today. Phil, let's start with you.
Mr. Hill:	OK. So, in general, for the policy standpoints I think it's important to develop a shared vocabulary. I think that's one of the themes that came out today and that some of the regulators have been grappling with. What is AI? What are the AI uses? What are the modalities?
	Like I said, you get taxed on what you don't know and so it's important for people to have these conversations, dig deeper into what people are using, with the Copyright Office and the Patent and Trademark Office taking comments from interested stakeholders, stakeholders participating and figuring out the way forward, because the Copyright Office and other regulators can only regulate about what they know.
	And so you have to bring the use cases, the business cases, the real stakes to their attention so that they can craft regulations that apply and actually matter in the marketplace. And then I'll say that there is in the United States maybe as the industry is moving very, very fast – some people may think it's a very slow process and, indeed, it is a slow process here – one of the things that people are trying to do here is ensure that all of the developments that we have are future proofed, that whatever we decide today doesn't hurt us 20 years from now, and so those are some of the interests at stake here.
Ms. Elluru:	I'll put it very simply but I know it's much more complicated than that. But I think we need regulations and laws that harness the benefits of our AI technologies and fuel our innovation so that it's serving humans and society. What that looks like is complicated because there's a lot of issues – IP, copyright, responsible AI. But I think that's what we should be striving towards.
Ms. Parker:	Yeah. I think, broadly speaking, I'm a technology optimist. I think technology can be used to solve so many tough challenges, and I don't want us to be so concerned about some of the potential harms that we end up sort of throwing the baby out with the bathwater.

And just one final little point that we haven't touched on today is making sure that as we think about these technologies and look at it kind of in an innovation forward perspective that we don't necessarily hold AI to a higher standard than we do the human-based systems.

I know a lot of conversation is around bias in AI but also some of these copyright issues are trying to hold AI to a standard of perfection. That means that we can't use it even though it probably would be better than a human system would be. So I think we have to keep that in mind as we try to come up with the right pathway forward.

Dr. Gupta: Yeah, that's a really good point. I mean, these issues are playing out in the real time. So I think maybe when we – and by the way, at CSIS with SCSP we've been doing regular roundtables on these specific topics. So if you are following it, you're interested, shoot us an email. We would love to hear from you, discuss with you.

But I think we have to come back. We have to come back to this topic. So with that, maybe we say thank you to our panelists here. We break for lunch and after lunch we'll reconvene at I think it's 1:00.

Audience Member: 1:15.

Dr. Gupta: 1:15, very good. Thank you. (Laughs.) And we will go straight into the future and problem solving world with some excellent panelists lined up. Thank you all. (Applause.)

(END.)