



Transcript

Exploring Generative AI and Law: ChatGPT, Midjourney, and Other Innovations

April 21, 2023

Contents

Welcome & Introduction.....	2
Panel: The Practice of Law and Large Language Model (LLM) AI Advances	15
Panel: Generative Art, Music & Intellectual Property Law	42
Panel: Potential Positive/Negative Impacts of LLMs on Governance & Society.....	62

Published May 16, 2023

Welcome & Introduction

<https://youtu.be/jjt78XoS8o8>

[00:00:01.08] HARRY SURDEN: Welcome to the Silicon Flatirons conference on Generative Artificial Intelligence and Law at the University of Colorado Law School. Delighted to have you here. I am Harry Surden, I'm Interim Executive Director of the Silicon Flatirons Center. I'm also Director of our Artificial Intelligence and Law Initiative, and also Associate Director at the Stanford Center for Legal Informatics or CodeX.

[00:00:31.11] So let me tell you a little bit about the Silicon Flatirons Center. This was founded by Professor, and now Attorney General Phil Weiser in 1999. And the goal of the Silicon Flatirons Center at the University of Colorado Law School is to initiate, sustain, and elevate the conversation about technology policy and entrepreneurship, and inspire, prepare, and integrate students into these really important areas.

[00:01:00.33] So we are so pleased to have you here at this conference. This is going to be a really important conference, I believe. We're at the cusp of something big and a fascinating day on generative AI and law. There have been tremendous advances in artificial intelligence in the last year, some of which I'm going to show you this morning and we will be talking about today.

[00:01:25.69] But before we get started, I want to have a whole bunch of thank you's for all the people who made this possible. First, I want to start with our traditional acknowledgment that the University of Colorado sits on land that is the traditional territories and ancestral homeland of the Arapaho, Cheyenne, and Ute people. So we want to acknowledge and be grateful for that.

[00:01:49.90] I also want to thank the tremendous Silicon Flatirons staff who worked so hard to make this conference come together, our amazing Silicon Flatirons team, including Nate Mariotti, Shannon Sturgeon, Christine McCloskey, Sara Schnittgrund, and many others. So please join me in thanking them.

[00:02:11.13] [APPLAUSE]

[00:02:15.15] Yes. They have been working around the clock to bring this together for the last couple of weeks. I want to thank all the Silicon Flatirons supporters and community who are here, many of which who have been with Silicon Flatirons for quite a long time. I'm really grateful to our incoming executive-- permanent Executive Director, Brad Bernthal, who could not be here today, as well as the Dean of the Law School, Lolita Buckner Inniss, who also could not be here today.

[00:02:44.73] I'm on a grateful to the law school faculty and staff and our faculty directors at Silicon Flatirons, many of whom are here today.

Our expert panelists for flying in far and wide to come. Our students, our tremendous students, many of whom are in the audience here. We're a very student-oriented organization, so I encourage you to talk to the students.

[00:03:10.08] Tell them about what you're doing out there in the world. They really are eager to learn from all of you. And we have amazing students here at the University of Colorado, so we're so grateful to have them. And many of them have volunteered to put this together. And finally, I want to thank all of you for coming, taking a big chunk of a busy day to come out here. And there are several hundred people joining online as well, so I'm grateful for them for joining us today. So thank you.

[00:03:40.11] All right. So let's talk about the agenda for today. I'm going to start with about a half-hour overview of generative artificial intelligence technology. I know that people are on different pages. Some of you are very advanced. Other of you might not know much about it, so we're going to try and bring you all up to speed, and that'll be a good foundation for some of our future topics.

[00:04:04.14] Our first panel will be on the use of generative AI, particularly things like GPT, or ChatGPT, or things built on top of that in the legal practice, and some of the implications for that. Our second panel will be very interesting. It's the implications of AI-generated art, music, video, and text on intellectual property law, things like copyright, patent, fair use.

[00:04:32.94] And then our final panel will be on the implications of this technology on society. This is a big deal, I believe, and what implications it will have on law, governance, social norms, things of that nature.

[00:04:50.93] So without further ado, let us begin. So as I said, there have been some tremendous advances in AI in the last year alone. I've been a software-- I was a software engineer before I was a law professor, so I have been studying this for over 20 years. And this is some of the-- probably the biggest advance in one year I've seen in the 20 years I've been studying this.

[00:05:16.78] Things are moving at an accelerated rate this past year that I've never seen before, so I think it's really important that we have this discussion now and get this on the table. So let's start with an overview of the technology to just get us up to speed. We'll talk about what is generative artificial intelligence.

[00:05:40.57] I'll give you an example of ChatGPT and GPT, and we'll understand the very basics of how it works, the capabilities, and the limits. We'll see how it's being used in law. And we'll just preview, but not really talk about the societal implications because our panels will be delving into that.

[00:06:02.77] OK. So let's just start with a basic question. What is generative AI? So that actually brings us to a more basic question, which is, what is AI or artificial intelligence? So there's no one definition that anyone will agree on about artificial intelligence, but one useful definition that I often use and I find helpful is the following. Artificial intelligence or AI is using computers to solve problems, make predictions, answer questions, or make automated decisions, or actions on tasks that when are done by people typically require intelligence.

[00:06:42.04] So there are a lot of different discussions about what intelligence is, and we won't get into that here. But just the idea is intelligence are a series of higher-order cognitive skills and processes, such as abstract reasoning, problem solving, learning, visual processing, language, et cetera.

[00:07:02.98] So when humans do tasks, such as driving a car, or playing chess, or writing text, these are thought to involve these higher-order cognitive skills. So when a computer does any of these tasks, even though they do it very differently, we call this an artificial intelligence task.

[00:07:21.98] OK. So what is generative AI? Generative AI is a subset of artificial intelligence tasks focused on generating outputs that are normally associated with human creativity. These are things like art, music, text, video. So AI can be used in lots of different ways, driving cars, playing chess.

[00:07:45.75] So one subset of this is generating human creative content. So if you'll notice the logo there-- I'm not an artist, but I generated that. Not because of my artistic skills, but we will see through AI software that helped me generate that. I could never have generated that on my own.

[00:08:07.92] So let's see an example of this. This is software known as Midjourney, and I just prompted it to create the Mona Lisa in the style of Pixar. So this is an incredible-- here, let's watch that again. Anyone can have availability to this. It's online. And you can basically just type in anything you want, and it will generate a version of that instantaneously, or within a matter of seconds.

[00:08:40.59] And I could never have produced this artistically. I'm not a great artist. But any of us now can generate something like this. So this is the generative art. There's companies like Midjourney, Dolly is another one, Stable Diffusion, there's a bunch of these out there. So it raises the question, who is doing the generating of the art? It's certainly not me. All I did was type that in, the AI did that.

[00:09:06.41] At the moment, it requires-- you can't exactly control what you're going to get out of these, and it requires some prompt engineering, they call it. I think in a year or two-- I think that's a

temporary phenomenon. In a year or two, you'll be able to much more accurately control what you're going to get out of this. Another example of this, a robot law professor teaching law school. So here we go, kind of a lifelike example of that.

[00:09:38.25] So this is amazing. This didn't exist. Two years ago you could not get extremely high-quality lifelike outputs. So that's generative arts. There's also generative AI in music. So this is one of many AI music platforms. So I generated this theme in seconds. You saw me do it just there.

[00:10:02.93] Any of you can do this as well. There's a bunch of sites like this where you can generate orchestral themes, or electronic music, or hip hop, whatever you want in any genre instantly, and totally created by the AI. Again, I'm not a musician, and I was able to create this. Totally unique.

[00:10:24.38] But while I think AI-generated music and art are interesting, that's not what I'm going to spend my time on today because I think, by far, the much more impactful technology here is AI-generated texts. These are large language models like GPT. And the reason I will show you why they're much more impactful because they have to do with information processing, knowledge generation, which is the foundation of everything else.

[00:10:51.87] So this is what is going to be the most disruptive, in my opinion, although the other technologies will be disruptive, for both better and for worse. When I say disruptive, I don't necessarily mean to imply bad as we will see, I think there are very positive disruptions that will happen, and negative disruptions, and our third panel will highlight that.

[00:11:14.39] OK. And in case there are questions, we will be having an opportunity after the first panel to have some questions. So please hold your questions until then.

[00:11:26.90] All right. So let's just dive right in. I want to bring you guys up to speed. There's a lot of terminology out there, GPT, ChatGPT, large language models, what does this all mean? So just starting out, ChatGPT is that chat-based interface to an actual underlying technology called GPT. So we're going to talk about GPT, but chat is just one of many ways you can interact with this underlying technology called GPT, which we'll talk about in a second.

[00:11:56.69] GPT was originally developed by a company named OpenAI as just a text generator, something that's capable of doing fiction writing, or story writing, but it actually turned out to be much more. But let's just see a silly example of what most of us-- let me just see, who has played with ChatGPT here in the audience? So yeah, so maybe like 80% of you.

[00:12:21.87] And most of us have used it for silly things like this. So I asked it to write a limerick about how great CU Law School-- University of Colorado Law School is. And in Boulder where mountains stand tall lies the law school revered by them all. It's pretty funny. It's clever. So most of us have used it in this humorous context.

[00:12:44.04] But it turns out it can do a lot more, a lot more. We're going to see. It turns out that later versions of GPT could do much more than just generating text, and this was surprising to everybody. So let's just take a deep dive.

[00:12:59.70] This is an example of GPT-4 writing a merger agreement between two companies. Something a lawyer or a business person would work at. And I've looked at all the outputs-- I've done over 300 experiments with GPT-4, and the quality of this, I would rate as a first draft level of a first-year associate, which is astonishing given what we saw in the past.

[00:13:29.01] Now, you wouldn't want to send this out in a deal right now, but you'd want to double check it. But the fact that it could produce a first-year associate at a law firm first draft quality in 30 seconds is absolutely astonishing to me.

[00:13:43.74] OK. So let's do an overview of GPT. What is that? It's a lot of jargon here. GPT stands for generative pretrained transformer, and we're going to break down each one of those words so we hopefully understand it.

[00:13:59.82] OK. So let's look at the word generative. Generative refers to the fact that it's part of generative AI, which I said is designed to produce outputs like texts, or movies, or music. In this case it's text, as opposed to using AI for some other purpose like playing chess.

[00:14:20.08] So our generative AI, if you look at this animation here, uses AI just to predict the next word that is coming along based upon what you've prompted it or asked it, and also based upon what it has produced so far in the text. So it looks back at what it's produced and what you asked it, and, amazingly, it is able to answer questions and do a lot of things that we'll look at in a second. But that's the word generative.

[00:14:51.48] The P is pretrained. So this is really interesting, and this is the secret sauce about-- one of the secret sources about how this works. So pretraining refers to the process of taking these AI algorithms, these transformer algorithms, and releasing them huge amounts of text.

[00:15:10.92] So GPT was exposed to a huge portion of the internet, over 2 million books, all the laws that were out there, and all sorts of other things. Their most recent version, they've been very secretive about what they released-- what they trained it on. We don't actually

know for GPT-4, which I'll talk about in a second, but huge amounts of texts.

[00:15:33.99] And then what it does in the pretraining-- this can take days or weeks, costs tens of millions of dollars, and it learns a couple of things. The first thing it learns in pretraining is how human language works in lots of different languages, whatever. So it learns the relationships between nouns and pronouns, words that come after each other, but amazingly we'll see, it also learns things like reasoning and other aspects that were unexpected.

[00:16:02.11] So because they are trained on billions of words, they are known as large language models. So LLM in a law school used to mean master of laws. These days, now, it means large language model. So that's sort of synonymous with the group of these technologies like GPT. But there are others I should mention.

[00:16:24.81] There's another very good technology from a company called Anthropic called Claude Plus, which is an advanced version of their Claude. Google has BERT. So I'm mostly going to focus on GPT, in particular GPT-4, but just note that there's a family of these large language technologies that more or less do the same thing, but some are much more capable than others. And in my opinion, by far the most capable is GPT-4, which we'll be focusing on today.

[00:16:57.54] The last word, T, is transformer. So what is a transformer? This was actually a special architecture for artificial intelligence neural networks developed by Google in 2017 in a famous paper called "Attention Is All You Need."

[00:17:14.19] And, more or less, it allows the AI to look at the context of what's been asked and what it's written, and make sense of that. So it does context aware interpretation, I'll talk a little bit more about this. But this is a huge breakthrough. This is one of the huge breakthroughs that allowed GPT to happen.

[00:17:37.20] So the transformer is an example of another piece of jargon called deep learning, which itself is in part of neural networks, another piece of jargon. I won't really be going into that, but I'm happy to talk about that. I gave a primer yesterday, which we'll talk about. If you want to see this in detail, I gave an understandable primer to law students that goes into all these technologies at some depth but in nontechnical terms.

[00:18:05.28] OK. So there's a couple of nuances that are really important to emphasize here. So, again, I said OpenAI developed GPT based upon Google's transformer architecture. So OpenAI borrowed this technology, and they made several crucial improvements, really important contributions. So they really made a bunch of big improvements that we'll talk about here.

[00:18:32.01] One is called instruction finetuning, the other one is called reinforcement learning from human feedback. Don't worry about that for now. I'll get that in a second. The most important thing I want to highlight on this slide is that there have been five versions of GPT. GPT-1 in 2019, eh, not that good. Sorry, 2018. GPT-2 in 2019, OK.

[00:18:57.83] Finally, GPT-3 in 2020 was pretty really good at generating text. But as we will see, it was not really good at doing other things like problem solving. The big advance we'll talk about here was ChatGPT, which most of us use in November of 2022. That was a huge breakthrough which involved instruction finetuning and reinforcement learning, two things we'll explain in a second.

[00:19:24.06] And then the biggest thing I want you to pay attention to is GPT-4, which came out just last month in March of 2023. This is the model that all of my examples are using, and it's really important not to confuse the free version with this version, which is only available to paid subscribers. So everything I show you here, you can replicate at home as long as you use GPT-4, and don't use the free older version, GPT-3.5. People get confused by this all the time.

[00:19:55.16] And then GPT-4 is a huge advance over the free version that most of you are using. So I just want to make that clear. And you can recognize that if you see the green icon for ChatGPT, it's the old 3.5. If you see the black icon, it's ChatGPT using GPT-4. So ChatGPT, if you pay for it, can access either of the most advanced GPT-4, or the older models, whereas the free version is only the older 3.5. That's a really important nuance that people miss.

[00:20:31.85] OK. So why am I so excited about this, and also a little quite scared about this? It's important to understand, as an AI researcher like myself, the huge leap that took place just last year from the beginning of 2022 to the end of 2022. This was a giant leap. So as I said, GPT-3, this text generator was pretty good when it was released at generating texts, lifelike stories, or press copy, and things like that.

[00:21:04.91] But they were not that good at doing other things like generating legal documents, or solving problems, or answering questions. In fact, they were pretty bad about that. So fortunately, we can travel back in time and see just how bad they were to get a comparison because there's still-- online there's a time machine of GPT because OpenAI keeps their older models accessible so you can see what things were like. So let's see how GPT was just last year, at the beginning of last year in January 2022.

[00:21:40.13] So I have a series of common sense questions that I've been asking AI models for years. Things that a toddler would know and laugh at. A toddler, if you asked it how many legs does an apple have, they would laugh and say, apples don't have legs. For years, the most advanced AI models could not solve this, nor could GPT-3 back in January 2022 just a year ago.

[00:22:05.13] I've been testing this for years. I asked it to write a legal document. For lawyers in the room, I asked it to write a summary judgment. That's not very good. The older GPT-3 model did not exactly appear to be a shining beacon of legal intellect, right? So you can see why many of us discounted this technology.

[00:22:28.96] But in November of '22, OpenAI released an advanced version of GPT-3 called ChatGPT, which they labeled 3.5 that had new engineering advances. And this is what we've all been playing with with the chat interface. So it's not just GPT-3. They actually improved it in a number of ways, so that's why it has the 3.5.

[00:22:54.22] And this is what shocked everyone, and this is why, I think, everyone, including myself was amazed. It turned out it was good at generating funny poems and things like that, but it was also able to engage in reasoning and problem solving. And that was 3.5. 4.0 is even better.

[00:23:13.22] So here's a puzzle here I gave GPT-4. It's not really important the content of this puzzle. It's more that I couldn't solve this puzzle. And it's doing an amazing job, and it got it right.

[00:23:27.94] And the general idea of this is that none of us-- none of our AI researcher expected a machine that generated text to be able to do automated reasoning and problem solving. It was a completely emergent property that was unexpected by people that study the space.

[00:23:46.36] And previous AI versions could not do this that well. They did a very bad job. So in one year, it went from, I'd say, to use a technical term, pretty bad to pretty good, and a shocking leap that most of us thought would take 5 to 10 years, who study the space.

[00:24:05.83] And we soon realized that it was not just a humorous text generator, it could be a general purpose tool for many things, including law. So here is it writing a patent application for a made up item that I call axolotlite. And I teach patent law. This is not a perfect patent application, but it definitely is good enough to pass first muster through a patent examiner who's not really busy-- who's busy and not paying attention.

[00:24:33.91] And this was amazing. A layperson could definitely not generate in a minute in the past a patent application that could pass muster at the patent office in 30 seconds. And now we can, so we're going to talk about this on the third panel. So I want to emphasize the giant changes that happened in one year.

[00:24:53.00] So remember, this is what January of 2022 GPT-3's best efforts were at generating a motion for summary judgment. Not exactly that great. This is GPT-4 generating a motion for summary judgment from just a random complaint I found on the internet from some

lawsuit. I just copied and pasted the words, plugged it in there, and look what it generates from that.

[00:25:20.08] I mean, just an astonishing-- for those who are nonlawyers in the room, that's really good. That's, again, the level that a first draft from a first-year associate. Now, you wouldn't want to turn this in, again, to the judge without double checking. But you wouldn't want to turn in a first-year associate's draft either without double checking. So it was able to cite the law, cite the relevant laws. I didn't feed it any of that.

[00:25:46.51] OK. What about the apple question? How is it doing about that? The old GPT-3 said an apple has four legs very confidently. Let's look at what GPT-4, the most advanced, says. How many legs does an apple have? OK. Well, that's pretty impressive. It's giving a detailed explanation. It knows that an apple doesn't have legs.

[00:26:10.77] All right. Let's push it a little bit further and start seeing if it has a world model. If it can figure out what happens with apples. So I asked it to-- what happens if a person takes an apple from a bag, a second person takes it from the bag, goes in the other room, and the first person hears chomping sounds, what will the first person think happened to the apple?

[00:26:33.68] So if you think about what's going on there, it has to get in the mind of the first person and figure out what chomping sounds is implying, and looking from the perspective of the first person. This is incredible. You'll see it figured out from the perspective of the first person what they would believe that the first person believes that the second person is eating the apple. That is astonishing. No technology was able to do this-- even close to this.

[00:27:01.76] I pushed it further. I asked it, if the apple belonged to person A who was intending to eat it, and then person B took it and ate it, what emotions would person A be feeling? And, of course, it got it. Person A would be annoyed. Again, that's astonishing. It had to do theory of mind. Put itself in the mind of the person, and figure out what the emotions were.

[00:27:24.28] Then I pushed it one more step. I said, imagine five minutes from this scenario, where is the most likely place this apple would be, right? And it figured it out. It's partially in the stomach of the person eating it, but it depends on how fast he's eating it.

[00:27:39.68] Some might be in the person's hand. Some it might be on the plate. Some might be in the-- so it had to have this imaginary scenario, put itself in the mind of these imaginary people, and then project forward in the future of this imaginary scenario taking into account what happened.

[00:27:56.03] So all of you can do stuff like this if you have access to GPT-4. Nothing I'm showing here is magic. You can all do this at home. So this is why I believe that GPT is one of the biggest breakthroughs in AI in the last 20 years, maybe more. And especially GPT-4, I want to emphasize. Not just ChatGPT, which already was good.

[00:28:17.69] So I'm just going to give you an overview of how GPT-4 works, but I'm not going to go into all of this thankfully. If you want to learn how in-depth it works, go to my primer, which is online. Yesterday I did a student primer. I'm just going to highlight a couple of things that we should pay attention to which I think are new.

[00:28:41.45] One thing to pay attention to is something called transfer learning. So this is something that's been a dream of AI for years. It never really worked. So the general idea is-- in the past, most AI systems were narrow. They would do one or two things. Identify cat pictures, or play chess, or do one or two things, specifically tailored to that output. But there was this idea like, could we create something that's more general that could transfer knowledge from one area to another?

[00:29:13.82] And they've basically done a version of this with GPT. So GPT-4 is trained on this giant corpus of words, billions of words. It learns these patterns. You can think actually of GPT-4 as having compressed a huge portion of the textual output of humanity numerically. That's mind blowing, but that GPT-3.5 has 175 billion parameters, they call it, which are the encodings of the text.

[00:29:45.71] GPT-4, we don't know, but it's rumored to be over a trillion. And that is really a compression of some portion of humanity's textual output, knowledge, and reasoning. Processes compressed in trillions of numbers, which is amazing. But this is what transfer learning - if you train a model that can learn language, generally, and learn reasoning, you can set it on lots of different tasks, including legal, what have you, medical. That's just the general point. This has been a dream for years of AI and hard to get.

[00:30:23.51] So one thing I want to point out, too, I forgot to mention, when I use words like understand and general, I use it metaphorically. I don't want to anthropomorphize this technology. It's still a machine. There's no evidence it's alive or sentient. I'm just using shortcuts. It doesn't really understand the way we do. It understands in a sense that it's responsive. Responds appropriately to the instructions, and previous models didn't respond appropriately to the instructions.

[00:30:54.87] So I just want to make that clear. So I'm using some shortcut language, but I don't mean to imply there's a little human living in there. Although it's basically a simulation machine. It's able to do extremely good simulations of humans, and that's what happened if you followed the AI Wall Street Journal reporter's interaction with Bing.

[00:31:17.56] So just a quick highlight. One of the big engineering advances that made this happen was something called instruction finetuning that GPT-- that OpenAI did. So apparently, we didn't realize this, but the GPT-3, which was producing the bad output had this capability lurking in it all along. It just took some clever engineering to extract it. So that's basically what OpenAI did through something called instruction finetuning.

[00:31:47.30] And so finetuning is where you take the general model, which knows about language and reasoning, and you teach it to be better at certain tasks. So they basically figured out that if you gave 15,000 input-output pairs of examples of good and bad answers--

[00:32:07.88] So, for example, write me a poem. And they had a human write a poem, and that's an example of a good poem. Or, here's a puzzle. Solve a puzzle. So they did this 15,000 times, and it turns out they were able to draw that latent capability from 3.0 to ChatGPT. So that was one of the big advances that made 3.0 come up to 3.5.

[00:32:31.81] The other one which is remarkable was called reinforcement learning from human feedback. I won't really go into details, but, basically, once they trained GPT-3.5 to know what good and bad answers look like, they then used this to train another AI system called a reward model by having humans rank outputs from best to worst, right?

[00:32:57.04] So if a human-- if the AI model produced a poem, it could produce five poems, then the humans would rank. This is the worst one. This is the second worst one. This is a good one. And it would learn, numerically, the difference between good and bad output. They trained a separate AI model called a reward model, and then they took that other AI model, and had it finetune ChatGPT automatically by saying, good output, bad output, good output, bad--

[00:33:24.82] So that was a real breakthrough because they no longer needed people to say what's a good or bad output. They could do this programmatically. And we'll see that's one of the big changes that's going on in AI is AI systems have learned to very effectively train other AI systems for better or for worse. And in the past, they weren't that good at that even though that was, at least, a possibility.

[00:33:48.16] Just a couple of examples of GPT in law. Again, these are experiments. I'm not advising that you do this. I'm just showing the capabilities. Here, this is an example of it doing legal reasoning. This is a really complicated evidence hypothetical that I couldn't solve. I asked it to solve it, and it answers it, and explains it's reasoning correctly-- it's amazing-- citing the correct area of law. So this is impressive.

[00:34:18.31] I teach IP law. So this is GPT-4 doing a fair use analysis. I'm doing something called chain of thought prompting, which I'll talk

about. But at any rate, I'm asking it to first list the process for performing fair use, which is when you can use copyrighted works without the permission of the author, usually, for some socially beneficial or transformative purpose.

[00:34:45.70] And then I asked it to compare Weird Al Yankovic's song, Eat It to Michael Jackson's song, Beat It, and it did an amazing job. So this is typically thought of as a hard problem to do because it's very subjective. So it did a nice job. But GPT is not perfect. Let me explain. There's a lot of problems. So I want to talk a lot about its limitations.

[00:35:13.49] It engages in what are known as hallucinations. So it sometimes makes up facts. So that's a problem in law, right? Facts in law are important. I will say that the most recent model, GPT-4, is a lot better. So ChatGPT, which is 3.5, is much worse about making up facts, and GPT-4 they've improved this a lot. But it still gets things wrong.

[00:35:38.66] And as I'll talk about, I think there is a clear path forward, as we'll hear about in the first panel, to improving and reducing hallucinations. So I think this is a problem of the current moment that can be reduced. So I don't think this is a forever problem. Maybe not eliminated, but we can get it better. It does get reasoning errors on it.

[00:36:01.21] So, again, it's not perfect on reasoning. It's not perfect. It sometimes gets common sense. It's just much better than what was possible in the past. That's the thing. But you still need to double check its reasoning. And that's a problem because it's so good that it appears that it gets everything right, but it doesn't, and that could lure us into a false sense of security.

[00:36:23.69] So here's an example where it missed nuances that my best law students would get. It's like at the level of an OK law student but not my best law student. So this is a nuance that only IP attorneys would get at where I asked it to help classify Netflix trademark. And the details are not important, but GPT-4 got it wrong, but then I gently corrected it and then it got it right. So it's open to correction.

[00:36:56.41] So Bing, as Microsoft's chat, has a version of GPT-4 that they've trained, but it's much more aggressive than OpenAI's version. And it wouldn't back down. Are you with me? And it was like, no, you are wrong. Netflix is a fanciful mark, which I think most IP professors would disagree with, but maybe they'll agree with Bing.

[00:37:22.12] But it wouldn't back down-- so take it for what it's worth-- even after I explained the reasoning. So you can gently explain the reasoning, and then OpenAI's GPT-4 will listen to your reasoning and correct itself for being whatnot.

[00:37:36.14] So there are ways to make large language models more accurate through prompt engineering. In our first panel we'll hear from

Pedro Arredondo, who has a startup called Casetext, which is using GPT-4 in the background. But there's something called retrieval augmentation.

[00:37:57.56] So you can either ask GPT to look in its 175 billion parameters and give you information, which is not that great because the model was frozen in 2021, or you can reach out in the real world to a search engine, grab relevant information, and then append it to your prompt, and then ask GPT to take that into account. And that results in much more accurate information. That's called retrieval or context augmentation, and that's been a successful strategy.

[00:38:28.19] Another thing is called chain of thought prompting. So this is remarkable. GPT can't see forward. It can only look backwards. It's only just predicting the next word based upon what it's predicted-- what it said previously, and the prompt you've given it, right? So one way to get it to be better as it can't plan is you ask it questions in stages.

[00:38:50.87] You say, hey, come up with a process for how you would do this, and give me that process. That's called chain of thought prompting. It's research by Google. Look back at the process you developed, and now answer it applying that process.

[00:39:04.61] So because it can't look forward and figure out the process, but it can look backwards and see the process it developed and apply that. So that's another way. Another thing you can do is, when it's wrong, ask it to reflect on what it got wrong and respond, and also verify, right?

[00:39:24.61] So I think the future-- so right now we're interacting programmatically directly with GPT through ChatGPT. I don't think that's the future. I think there's going to be an interface like Pablo's company Casetext in between the user and the model that does all this smart stuff, so the user doesn't have to do it.

[00:39:46.12] It gets the right information, it verifies it, before it gives it to the user. I think we're just in a moment where this is early technology, and right now we're interacting with the model directly. But I think in a year or two, some of these problems will go away.

[00:40:04.61] And so I don't necessarily recommend that lawyers use this directly, these are experiments, I want to emphasize. Definitely don't put private client data or privileged data in there. You could lose privilege. You could violate ethics laws. But if you don't use private data, you can use it to generate ideas as our panel will talk about. There are some uses for it.

[00:40:32.57] There are a bunch of limitations that are beyond the scope of this talk, but I want to raise them because they're real. And we can talk about them in the Q&A, and they shouldn't be disregarded.

But I do want to talk about this technology will be disruptive for better or for good. So there will be a lot of positive disruptions that are access to justice, increased efficiency.

[00:40:58.97] And, also, negative ones, universities are going to have to adapt, law schools, government, lawyers. But I think it's-- we don't really know what's going to have to happen. So I think there's a lot we need to learn, and we'll learn from our terrific panelists.

[00:41:17.78] So speaking of our terrific panelists, in a minute we're going to bring up our first panel, where we're going to talk about the practice of law, ChatGPT, and other AI advances. Well, we'll hear people who are on the ground using large language models in law. So I think it's amazing for better or for worse.

[00:41:37.67] But before we get there, let's look at a review that ChatGPT wrote of this conference three days before it happened. So I asked it to write a positive review, and then I just cut and paste the website, and so let's see what it said.

[00:41:57.29] Wow, an unforgettable experience. The organizers left no stone unturned. This is amazing. So I can only hope that we live up to that ChatGPT's expectations with this conference. Thank you very much.

[00:42:14.24] [APPLAUSE]

Panel: The Practice of Law and Large Language Model (LLM) AI Advances

<https://youtu.be/5GmVqymalcw>

[00:00:01.51] HARRY SURDEN: So thank you, guys. I'm the person who was over there a moment ago. I'll be moderating this.

[00:00:09.43] Unfortunately, or fortunately, you'll be seeing a lot of me because I am moderating the first and third panels. But we are absolutely delighted to have a terrific expert panel here talking about the use of large language models in law. Our panelists are extremely accomplished.

[00:00:26.23] In the interest of time, I will only give a very short biography of our panelists, but I encourage you to look up their very extensive accomplishments online. Let me start with Pablo Arredondo. He's Co-founder and Chief Innovation Officer at Casetext, which is one of the first companies using a product called Co-Counsel, which actually is a version of what I was suggesting as a good idea in terms of putting a layer between the user and the direct GPT model, so he--

[00:01:01.71] And Pablo can talk about it. His model is for lawyers. And on the back end, it works with GPT-4, but make sure that the lawyers

have their data safe and secure, so they're not just uploading private information to OpenAI.

[00:01:19.37] Next, we have Megan Ma, who works with me as Assistant Director of the Stanford Center for Legal Informatics, or the CodeX Center at Stanford University. Megan has a PhD in law and linguistics and is one of the leading experts in AI and law. Next to Megan is Daniel Schwartz, who is a law professor at the University of Minnesota and has done really interesting work on GPT and its use in law.

[00:01:50.35] And also, he did write a really interesting paper with several co-authors where they tried to trick law professors into seeing if they could determine whether an exam-written answer or written by GPT was as good as the student's. And he could tell you-- not really trick but see if they could figure out and it could perform at the same level. I won't give the spoiler on that.

[00:02:16.91] And last but not least is Jason Adaska, who has been working for years in AI and law and is the Director of Innovation at Holland & Hart LLP. So thank you to this amazing panel for joining us. I'm going to start out with the first question to Pablo.

[00:02:36.79] So Pablo, what are some of the abilities that this technology, and we're talking about in particular GPT-4 and large language technologies, can do compared to just early last year and before?

[00:02:50.80] PABLO ARREDONDO: All right, yeah. Thanks. First of all, thank you guys so much for having us. This is a real privilege to get to talk to you about this stuff.

[00:02:58.00] So we were shown GPT-4 very early in September of last year. And my co-founder, Jake, and I, basically, within 48 hours, had pivoted the entire company to do nothing but focus on this. We had been working with large language models since their inception five years earlier. And we had been shown GPT-3, which we thought was neat and had some cool tricks but wasn't ready for prime time.

[00:03:21.22] What we saw in GPT-4 was basically a literacy that was of a qualitatively different nature than anything we had seen. And while you're going to hear this called the generative AI revolution, I submit to you that for the legal profession it's really not that they can generate text that matters, it's that it can read it and interpret it and annotate it, and structure it and restructure it. So for example, we have a Fortune 50 client that was a beta client and said, we have these little nemeses, these little expert witnesses that just wake up every day and testify about how our products aren't safe. That's how they make their living. Could I take every expert report that this guy has ever done and every deposition transcript they've ever done and give me questions for cross-examination, finding inconsistencies between what they said?

[00:04:10.45] And I said, look, we'll try this, but I just don't think that's-- come on. That's a bit much. It's not going to work.

[00:04:15.94] Well, I was wrong. I said it was not going to work twice. It worked again for somebody else.

[00:04:19.96] So that ability to go through and substantively identify inconsistencies in a way that an attorney would find those useful things, that is not grunge work. That is not just putting on certain forms, filling out forms in a certain way. To my mind, that's the thing that we see GPT-4 doing, whereas even GPT-3.5 really just wasn't up for it at the same level.

[00:04:44.68] And did you mention the bar exam? Have we done that?

[00:04:47.02] HARRY SURDEN: No, go ahead.

[00:04:48.05] PABLO ARREDONDO: All right. So we were working with OpenAI. And I was like-- our colleagues, Dan Katz and Michael Bommarito, had used the earlier model and put out a paper called "GPT Takes the Bar."

[00:04:58.12] They should have called it "GPT Fails the Bar" because a lot of confusion happened from that. But the earlier model failed miserably. It got in like the 10th percentile.

[00:05:06.03] Well, with GPT-4, we got in the 90th percentile. We redid the study. And then, for good measure, we actually included essays and the Multistate Performance Test to write the full bar exam.

[00:05:16.36] So I think I'll end with this. ChatGPT is great for raising awareness. But to my mind, it's a little bit like, imagine a society that had never seen cars. And then here comes the first car, but everyone's just doing donuts on the lawn and then doing 90 and reverse on the freeway.

[00:05:35.37] I'm glad you know about cars, but using them well and responsibly is a very different experience than using them incorrectly, where you're typing in things and getting hallucinated at and all these things.

[00:05:47.07] HARRY SURDEN: Well, that's a really good point. One point I really liked about what you mentioned was its ability to read and synthesize as, something I didn't emphasize, which is just generating documents. And I really agree with you. That is a huge game-changer in law.

[00:06:03.84] Let me throw out the same question to the other panelists. What do you see are some of the new abilities of this technology? And just put up a finger if you're interested in replying. Yes, Jason?

[00:06:16.10] JASON ADASKA: Yeah, so continuing along the same theme, we've seen older models be very good at summarization, those texts where you take some input, a lot of different text, and needed to provide the highlights associated with it. But the inference, the reasoning capabilities, I think is unexpected not only to users but to my understanding to a lot of the researchers for this as well. That's the kind of delta where it's not just raw tasks or not just sort of document generation, but actually doing more complicated inference and legal reasoning.

[00:06:55.43] It's surprising because it is not the thing that you would expect a language model, something that just has patterns and language to be able to do. It's emergent just from seeing lots of examples. So I think there are questions that we have, which are, essentially, what are the boundaries? How complicated of a scenario can it get? And I think people who are doing experiments right now are trying to understand where that frontier is and how much it'll change in the future.

[00:07:26.21] DANIEL SCHWARTZ: The other thing I think is pretty important to understand is, so far, we've been talking in a sense about you ask the question, you get the answer. And then that's sort of the answer. And some of them are good, some of them not.

[00:07:35.69] What's really pretty amazing about it, in my mind, is that you can have this dialogue with it, where you get it to further refine its answers to match what you want. So a lot of times, its first answer may not focus on what you want, or may answer a different question, or you may actually realize your own question was not great. And this happens a lot in law practice.

[00:07:57.95] For folks who are an experienced managers, you'll tell an associate, hey, go write a memo about this issue or this. And they'll write it, and you'll realize either they didn't quite understand what you wanted, or maybe you didn't-- you weren't as clear as you probably should have been about exactly the scope of what you wanted. But because of how quickly it works, and because it retains a memory effectively, about that dialogue, you can sort of in real-time get it to adjust to the point where you want it.

[00:08:28.02] And I think that's something that folks who have not actually really tried in a sustained way to use this have realized. A lot of times, you'll see, Oh, it produced the first version of this, and it's not exactly what I wanted. It's not this is not good, or it hallucinated it. But if you sort of stick with it and use it in a sustained way, you can get much, much better very quickly at, say, having it draft a contract where you can say, OK, we'll draft that first version of contract.

[00:08:57.11] Gee, now find ambiguities in what you wrote. Great. Now, please expand on that one provision and have it create some incentive structure.

[00:09:06.98] So it really actually can, essentially, replicate what can sometimes be, in my experience, months-long process, where you get a work product back, you say go back and fix it, go back and figure out this answer. And you can't anticipate always where the issues are going to be. You can do that in real time in a matter of minutes.

[00:09:28.34] And so I think that's a really important element of the technology that folks need to work on because it's part of-- it is also part of-- it uses legal skills to be able to realize where the deficiency is, where do I want you to expand, where did you not take this and write in quite the right direction I had hoped.

[00:09:48.44] HARRY SURDEN: That is a really good point. I think, at least right now, the technology you-- it takes some getting used to figure out how to best use it.

[00:09:57.75] So at the beginning, I was not very good at it, and then you learn the things that it can do, and what it can't do, and how to work with it. And I think you're absolutely right to encourage people to not just look at the first output but to experiment with nudging it down the line, which is also one of the huge advances, as I said earlier, that we talk about natural language, quote, "understanding," because, again, it's not a little human in there. But previous technology could not reliably understand what you were asking or correcting it.

[00:10:34.10] We've all used these chatbots online. And you ask it, give me a customer service representative. And it says, do you want to order a pizza?

[00:10:42.05] So now GPT-4 very reliably understands exactly what you want it to. Meghan?

[00:10:51.23] MEGAN MA: I also want to maybe put out an alternative perspective that these advances don't also come from nowhere, as you pointed out. I think if you look into areas of cognitive science and linguistics, there are these traces of an ancestry where this isn't unforeseen entirely. There's an area called cognitive pragmatics. And within a subfield of linguistics, you see pragmatics is almost this contextual understanding.

[00:11:20.07] And you see this area that emerged about what you perceive as a conversation game between two humans. And part of this nudging or being able to tease out information-- there are existing sort of techniques that are done in linguistics to better understand and interpret one another. And we see that actually, with training on human feedback, that might have been one of the accelerators that, as you rightfully pointed out.

[00:11:44.49] So while these are kind of exciting advances, I think what's also exciting about GPT, generative AI, and large language models broadly is that a lot of these fields that were in disparate silos

are now coming into a deep intersection. And I think that's particularly what's making this especially interesting.

[00:12:02.88] HARRY SURDEN: Could you say a little bit more about the connection between instruction fine-tuning and other disciplines? Because I didn't know that. I think that's really interesting.

[00:12:10.95] MEGAN MA: Yeah. So this is a paper that predated just slightly the emergence of GPT. And there was a cognitive science and a programmer, his name is Evan Pugh, who looked into kind of the way in which human and machines communicate. And he basically asked a general question, why is it so unintuitive for us to speak to machines?

[00:12:36.21] And over time, they discover that it is because in the way that we communicate and task things between humans. We actually set out goals. And then, it's a search strategy to identify the solution to that goal.

[00:12:49.60] And so, they started to mirror or find ways in which they called natural programs. And I think that these techniques actually had helped. And it came in the form of, say, instruction fine-tuning.

[00:13:00.31] But essentially, what he did or-- and his team did was they tried to train a dataset and build a data set that is entirely built on natural language instruction between one another on abstract and reasoning tasks. And I feel like that really played a role in the way that ChatGPT and other large language models are now coming to be.

[00:13:23.40] HARRY SURDEN: That's fascinating. And it's also a plug for interdisciplinarity and working together with-- outside of academic silos. Any other comments on the first question?

[00:13:36.10] All right, the next question for Jason, so as we said, this technology is good, but it's not perfect. And we want to make sure we are very clear about the limitations. And so, as Jason, what do you see as some of the limits? And what do you see as short-term limits that will go away in the coming years and longer-term limits for which we don't really know what to do?

[00:14:00.54] JASON ADASKA: Yeah, no, it's a great question. So I think there are several limitations that people are thinking about, some which I think are short term and some which are going to be the longer term. The one that I think is on the top of everybody's mind, and it's one that you had mentioned at the beginning, was inaccurate or outdated information.

[00:14:20.25] The way most people are interacting with these tools now is essentially just asking it a question out of the blue, not providing any context. And it, with the newer models, is a lot better at being able to not hallucinate, but it's still it still can happen. And that's fundamental to how these models work. They're probabilistic pattern matching.

[00:14:44.46] There are a number of techniques that can be used right now to reduce that. One of those areas you had mentioned, which is not ask the system just to draw from its general knowledge, all of those 175 billion weights, what is the right case law, but actually provides some context and some sets of documents for it to reason about as part of the prompt. That's one technique right now that's extremely effective for being able to reduce hallucinations.

[00:15:16.26] There's also things that are happening in the ecosystem that are extending what these tools do natively by being able to pull in external sources of information. So OpenAI has released a beta version of what they call plugins. So you can ask it a question and have it not just use what you asked what was on trained weights, but actually provide connections to real-time data streams. It will go out, and it'll query and add to the prompt to get a lot better answers.

[00:15:51.31] So I think the verification, the up-to-date information, that's a problem right now, but I think just really in the next few months, is probably going to be less of a concern that people have. The other limitation, and this one, this one's a little bit amusing to me as somebody who's been working in technology in AI for a while, is these large language models are, in a lot of ways, the dual of what computers traditionally have been very good at. We think about calculators and computers as being things that crunch numbers, can do complicated decision tree logic if then. And large language models are based on these probabilistic reasoning.

[00:16:35.04] So for instance, you may ask it to do some reasoning about tax law, but you probably want to double-check the arithmetic that it's actually using to do calculations. Again, those are the things that I think in maybe the next year or so are going to be resolved not by any fundamental change to how large language models work but by incorporating other modules and plugins. So instead of just having the large language model, having to answer arithmetic questions itself, being able to use other pieces to be able to solve those.

[00:17:13.22] Folks have probably seen it play chess. It does chess relatively well, but it's not going to play chess as well as the Stockfish or some of these other custom systems that are out there. The third thing, and this is actually--

[00:17:30.14] At first blush, I think it may seem a bit pedestrian. But I think it's actually a relatively important limit for a lot of legal applications. And that is something that's going to sound silly, which is a buffer size associated with these.

[00:17:45.45] So GPT-4 can have as input and output something like 100 pages worth of documents and pages worth of words. There are a number of use cases that are specifically relevant for legal, where having the system be able to kind reason about a large swath of input and be able to make connections between some of those are going to

be important, and at least right now, is somewhat limited. There's engineering workarounds. But in general, it's going to live.

[00:18:19.17] I'll give you an example that our group has worked with. So if you are doing-- responding to something like an obviousness rejection for a patent application, the examiner responds back to you and says, I'm not going to give you a patent. I'm going to reject some certain claims because I've seen that there's a patent A and a patent B that's out there. And a clever person could combine patent A and patent B to describe what you're claiming is a new invention.

[00:18:52.05] An attorney who is responding to that now has to do reasoning, where you're looking across-- it's a triangulated information about a whole patent-- one patent, another patent, your own application and drawing inferences across. That's the kind of legal reasoning that requires a big, big working buffer of space.

[00:19:15.56] And right now, there are some limitations. There's some hard limitations for how many tokens or words you can put into this. There's also some question of how well it's going to work to scale.

[00:19:27.38] I think the way we're used to thinking about computers and technology is having hard limits measured in things like how fast the CPU goes, how much working memory do you have. I think this buffer size is something that it's going to be the next scaling parameter that the engineers are working are going to have to actually work hard to be able to expand. The four thing-- and this is not so much of a technical item, but I'll mention it because it's, I think, relevant for a lot of people who are putting the technology to use, as Harry pointed out law is one of those things where invention of facts is frowned upon. So I think that there has been two threads of conversation about using this technology in law, one which is, Oh, my gosh, this is amazing. It can do legal reasoning. It can actually do substantive work.

[00:20:27.32] What is this going to do in terms of supercharging the practice of law? And the other thread, which is, Oh, my gosh, it might make stuff up. And we need to stay away from this. This radioactive.

[00:20:40.38] I think the conversation about how to use the technology in a trustful way is right now-- it's not a technology problem as much as just a cultural problem for people to understanding how to use it, what are the right safeguards. I actually think that is going to take probably a couple of years for people to get used to, for clients to become comfortable with, for attorneys to become comfortable with.

[00:21:06.12] HARRY SURDEN: Those are really great observation. So thank you. Let me turn this out to the rest of the panel. Anyone want to talk about limits? Megan.

[00:21:13.87] MEGAN MA: So I think one point that you made, Jason, that I thought was really interesting is the idea of plugins and different

ways that folks out there in this field are trying to almost mediate for some of these limitations. And one paper that came out recently that I thought was really interesting was actually hugging GPT where, essentially, they were trying to leverage the strengths of ChatGPT, it being so great at communicating with humans, and then using it almost to triage to models that are built for specific tasks. And I think that we see a future where we don't necessarily have to have one model do everything.

[00:21:51.82] Yes, we see Midjourney, for example, as a particularly great example at text to image generation. We're going to see, I think, more and more of almost models becoming tour guides and directing you to what you want to--

[00:22:06.49] HARRY SURDEN: Could you just say a little bit more about what does it mean for GPT-4 as a model to talk to other models for those who might not understand that terminology or those ideas?

[00:22:16.45] MEGAN MA: Yes. So we're starting to see, almost part of the emergent behaviors between models is their ability to signal in a way various, I guess, tasks and to direct and say, hey, we think that-- this is a particular task.

[00:22:39.85] And they parcel it out and allow models that are-- so for example, Hugging GPT, basically, they had leveraged the fact that Hugging Face, which is this big repository of machine learning models-- they know that some models are better than others. And they use it as a segue into others.

[00:23:01.15] And to be honest, I'm not 100% sure of the technical elements behind that. But what I do see is this ability to almost build layers on top of models that are able to better refine what are the tasks and what are the work that is specialized for a particular field.

[00:23:22.45] HARRY SURDEN: Yeah, that's great. And then maybe a way to think about it, something Jason was saying, where the GPT is the middle person who listens to what's coming in and then decides, Oh, this is a math problem. I'm going to send it out to a calculator. This is an image generation problem. I'm going to send it out to an image generator. So it's the middle.

[00:23:43.27] PABLO ARREDONDO: I just want to elaborate a one that may have come up earlier, but also the scarcity of the chips to run this stuff. So Paul Lomio, who was the director of Stanford Law Library, when I was at law school, told me that in the early days of online research, they would tell the Stanford kids, you're not allowed to use it between 11:00 and 2:00 because that's peak New York time. So basically, you guys can't do online research. It's New York's turn to do it.

[00:24:05.32] And what I've found is that we find ourselves in something similar right now. We are literally burning through these servers that we have. There's more people wanting to use it than we can. And it's slowing things down.

[00:24:16.63] We're getting more servers. They're not only extremely expensive, but they're also just-- it's not. You can't just order as many as you want. They're like partitioning them out.

[00:24:25.37] And so Casetext, we've always been, for instance, all students could sign up for free, all judges, et cetera. We've had to change that right now. We've had to be much more judicious on who we can get it out.

[00:24:35.90] And so I think there's going to be some distributive justice issues with this, not the-- Westlaw brings its own distributive justice issues, just fine without AI. But I think we're going to see this as something that's pronounced that getting access to the actual chips to run, especially the really good latest models. I think the other models are getting better and faster and cheaper.

[00:24:57.61] HARRY SURDEN: Let me just follow up on that because that's a great point. Do you see that as a problem of the moment? So right now you can run models like LLaMA that aren't nearly as good as GPT-4, but two years ago would have required a data center to run, and now, you can run it on your laptop. Two years from now, do you see running something like--

[00:25:16.12] PABLO ARREDONDO: Yeah, there's so much capitalist pressure and evolutionary pressure. And, boy, that capitalistic pressure could do a lot. So I think there's a lot of incentives.

[00:25:25.66] But creating new plants to create these chips is not an overnight thing. This is something where it's just-- there's a lag time even when you decide to do it.

[00:25:34.64] And Microsoft is trying to corner the market on some. And then, is this just a complete geopolitical thing? But in the meantime, literally--

[00:25:42.97] And right now, we're like, oh, on demo, look how cool, yeah. It's not even working. It's so slow because everyone wants it.

[00:25:47.98] And it's working for right now, but I think that's going to get old pretty quickly when they start to--

[00:25:51.59] HARRY SURDEN: Fascinate. Dan, did you have comment?

[00:25:54.16] DANIEL SCHWARTZ: Yeah, I just wanted to follow up on another element of the-- because I think hallucinations answers the question of, is the AI making up facts? That's one of the biggest questions that are out there and biggest concerns.

[00:26:05.35] And one of the things, and we've talked-- and Harry talked about a number of techniques that can be used to mitigate that and how the eyes are getting better. But I think one the real possibilities here is using the AIs to help you fact-check.

[00:26:21.26] So at the end of the day, these AIs are good enough that you can ask them, look, substantiate your claims, show me the underlying text, give me the information. And so, actually, I'm one of the privileged few who has been able to use Casetext. And it has these amazing technologies where it doesn't just give you an answer. It will then give you the quotations from the underlying documents.

[00:26:42.65] So that it makes actually site checking, and the type of side checking you might do as a young associate or as a law review editor, relatively easy. And you can go back and verify it. You can even do this within GPT-4 now if you're--

[00:26:56.33] It takes a little bit more engineering. But you can say, look, please give me an answer. And then the next questions is, OK, provide me with a direct quotation from the underlying source that I can see to substantiate that. And then you can go back and do your--

[00:27:10.19] So there are tools in place to facilitate the type of site checking, the type of verification that you need to do that are, really, you'd want to do again. You'd want to do that with any legal document to make sure that the underlying references to the cases are accurate, the underlying references to the whatever it is the depositions, or the underlying emails are accurate. I think that it is much less of a hurdle than initially some people may have made it out to be that these systems hallucinate because you can also actually use, these systems to ensure that what they are saying is represented accurately in the underlying documents that you're providing it.

[00:27:57.22] HARRY SURDEN: Yeah, thank you for that. It's a terrific point. Let me put the next question to Megan.

[00:28:03.67] So many of us here are lawyers. And we're interested. Wow, can this technology be used in law? Should it be used in law? So Megan, where do you see these technologies being usefully deployed within law? And what are some of the benefits and risks?

[00:28:19.80] MEGAN MA: Yes. So I think how I see this question and put differently is what are the relevant use cases? And more importantly, what do I really have to do and in terms of changing my own processes to accommodate for using these tools?

[00:28:34.92] And I think just by the examples that you've shown, it's basically showing remarkable performance. But I think the operative word of that question really is "usefully," because a lot of what people think about is even when there's any technology out, it's like, will I really use it? That's why I defer back to Word, that kind of mentality.

[00:28:56.47] And I think, essentially, what we want to embrace in the coming future is what the possibility of having it leverage this type of technology to do basic legal work, which will enable better access to legal services, for example, or expansion into pro-bono services. We think of it as being able to help with legal aid clinics, for example, a separate issue in legal diagnoses, and helping to kind of service more clients in need. The other side of it is, because a lot of what we've contextualize right now is it's experiments, we're doing experiments, we have lots of experiments, we actually haven't really thought about integration into or practical integration into our processes. And so until we get past experimenting, that's when we really can move into what we think is practical use.

[00:29:49.35] And so some of the questions that we might be thinking of asking is, yes, it's capable of drafting contracts, but what then are the edits that our associates or counsel will have to do on top of that? We might see that they're conducting really complex legal analyzes. But how should lawyers then react to this type of analysis? What do we do further? What is that next step?

[00:30:10.62] And we can re-imagine, for example, new methods of IP being able to build legal arguments, for example, or maybe it's creating that first draft, that first template, and it also offers these very specific, highly specified annotated commentary that it took maybe a year or two for a first-year associate that's entering your law firm to then be able to pick up what experience means in your law firm. Now, you're able to do that through these annotations within your drafts. I think part of our center, one of the ongoing research elements that we work on, is trying to uncover how legal expertise actually differs across seniority and specialization. And so we're trying to better represent what actually is that value add that you get from seniority and partner level expertise.

[00:30:59.07] And what we're seeing here is these models being able to maybe capture these differences in legal opinions and use them actually as a strength. We're allowing lawyers to be able to gain new insights or expand their critical thinking. And going forward from that, we anticipate more of an embodiment. And what I mean by that is we're getting into a space where we can simulate circumstances of potentially negotiation, litigation, or merger strategy before it happens and other dynamic interactions that we weren't able to gain play before. Or for example, we had a sparring partner within the law firm, but now imagine that kind of as a crowdsource type thing through these tools.

[00:31:40.26] But to the second part of the question, there's no technology or tool that we use without risk. But one of the main issues, at least I see it, is that risk is ill-defined in itself, especially in the field of artificial intelligence and in large language models. So we heard questions about data privacy.

[00:31:59.52] We know that, for example, ChatGPT is being investigated in Spain. It's also being investigated in Canada. In Italy, it's been full-out banned. There's also a lack of transparency around the data that it has been trained on.

[00:32:11.76] We know that it's been trained on a lot of texts, but what texts exactly? And what are the weights? We don't really know.

[00:32:18.06] And there's also this "no man's land around" the protections of using these models. So I use these models. What do I do from there? This is particularly concerning, of course, when the information is sensitive, of confidential nature.

[00:32:34.12] And we've seen, of course, that Casetext with Co-Council has put in those guardrails. And so these concerns aren't necessarily there, maybe for those who are in a corporate, large-scale law firm setting. But what we're seeing is even between ChatGPT, this free version, and GPT-4, there are monumental differences.

[00:32:55.06] And so, the concerns around data remain at large than for the everyday person. And so the point I made earlier about access to legal services and leveraging these large language models to minimize this gap actually resurface if we have very large gaps between the performance of these models, the free version versus the paid-for version.

[00:33:21.00] And this is just a risk at the foundational technical level. Risks also can be looked at from the lens of use and interactions with these models and the harms that can come out of it. There's actually really well-thought-out taxonomies of harm and risk that are being put out there, actually, by DeepMind, Google themselves, and by communities of responsible AI and AI ethics communities. But they kind of remain at a level of generality. They don't translate well into a specific domain, such as law.

[00:33:50.59] And the question even about evaluations or auditing, you might hear from Anthropic that they put themselves out there as robust, safe, transparent AI, but what does that really mean? We don't even have a consensus around what are the relevant auditing tools or what evaluations we can benchmark against. And so, having these limited understandings, I want to think about the tools that we use that are most pervasive in our everyday work.

[00:34:17.35] We think about Google Workspace. We think about Microsoft 365. Actually, these companies are going to be integrating directly large language models.

[00:34:26.25] And so, already right now, in our practices, we have things like auto-complete grammar spellcheck. But if you think about it, that added layer on top, we talk about personalities. So Harry pointed

out the difference between even Bing and OpenAI. And it is kind of aggressive nature.

[00:34:45.75] We don't necessarily know what those personalities and how we're able to necessarily negotiate and speak back with those machines. And so that's probably one thing that is a risk that we need to be a little mindful about because harms don't actually always come in the place of being glaringly obvious. Sometimes they come in very subtle, nuanced, and behavioral nudges, such as those.

[00:35:07.45] So if you were not someone who knew actually very clearly that difference in IP questions and had that argument with Bing, would you succumb to what Bing has answered, or would you be able to negotiate back? So I think a lot more research needs to be done on actually questions of contextualized harm and risk. And I think that is what we'll have to do in this buffer period as we look more into a large language model.

[00:35:34.32] HARRY SURDEN: Wow. Thank you for that really comprehensive answer. And to your point, I felt like I'd hit a new life flow when I got an extended argument with an AI chatbot. So that was not my proudest moment.

[00:35:46.17] Let me toss this out to the panel. What do you guys think in terms of benefits, risks? Jason?

[00:35:54.72] JASON ADASKA: Yeah, so I guess in terms of incorporating this specifically into legal, one of the things that I think is really interesting about how this technology has caught on, and it's a little bit of a maybe differs in some ways from the point that you made earlier in your presentation here, Harry, which is, in the future, there's probably going to be other technology that sits in front, and it may not be a chat interface, I'd actually push back on that a bit. It's certainly going to be the case. This is going to be in a lot of tools.

[00:36:28.66] But I think one of the things that has allowed ChatGPT to be so successful is that it's in an interface that-- people don't have to learn how it works. You're using natural language.

[00:36:42.45] In fact, even the workflow for it drafts something, "no, that's not right. Can you please fix this particular paragraph?" That's the way the attorneys work now. The change is just that you're not working with a person in many cases. Now you're working with a machine that happens to be interacting in that same way.

[00:37:02.22] I think that what's exciting is when you have technology that the people who are using it don't have to change what they're doing-- it's the technology has come to them. And I think that's what we're seeing with the chat interface and large language models, is you have this very general-purpose Swiss Army Knife interface.

[00:37:23.68] You don't have to learn it. You don't have to read a user manual. You don't have to know what button to press. You just chat with it.

[00:37:30.72] And there's obviously, as we talked about, ways to do that more or less effectively. But they're really around what you would use to talk to a person or less effectively. I think there's always going to be that. And that's actually one of the things that will allow this technology to be really transformative, specifically in legal, which has a history of being pretty conservative in terms of, hey, let's change how we're working, let's use some other tools.

[00:37:58.23] I think I think people-- since the tool allows the interaction that are natural in many ways, I think that's what's going to help it get adoption.

[00:38:08.40] HARRY SURDEN: Terrific point, and friendly refinement accepted.

[00:38:11.72] [LAUGHTER]

[00:38:12.84] Pablo?

[00:38:13.32] PABLO ARREDONDO: So I oscillate between optimism and pessimism with this stuff as the optimistic side to me. So when the computer first came out, Isaac Asimov wrote an article called "Who's Afraid of the Computer?"

[00:38:27.33] And he opens by talking about Kepler and saying Kepler had these great insights into how planets move. And then, he had to spend eight months doing these tedious calculations. And can you only imagine what Kepler might have thought of if he had been freed from that tedious labor and could have spent those eight months kind of like shower thinking?

[00:38:48.21] So the optimist to me says, can you imagine if we put all of this grunge work down, all of this unnecessary, tedious, repetitive, non-intellectual aspect of law and let ourselves return to like the stuff we learned in law school to think about? What are the policy reasons underlying this? Having time to go find a Chicago sociology study that shows that the actual predictions, these things, this deeper advocacy that might be possible.

[00:39:13.48] So that, to me, is the good outcome. The bad outcome is this race-to-the-bottom, McDonald's-ization of the entire field, where it's all just cookie cut it out. It's good enough. And we lose some of the artistry of it. And it might sound strange to hear me talking about this as the guy from Silicon Valley who's selling these wares. But--

[00:39:34.50] [LAUGHTER]

[00:39:35.67] On the contrary, I was a-- I'm a lawyer. I still pay my bar dues. I think the legal profession, though much of it is, unfortunately, a

shadow of its earlier self, still has a lot of nobility to it. And so I would like to-- I hope that we can use this stuff correctly, to then allow us to both give more people representation and to really increase the quality of representation. That's a good outcome. The bad outcome is just, yeah-- what I described a little.

[00:40:03.66] HARRY SURDEN: Yeah, great point. And just to clarify, we don't-- you don't sponsor Silicon Flatirons at all. And we are not, but you're one of the first to-- in the space. That's why we're having it, yes.

[00:40:15.03] [INTERPOSING VOICES]

[00:40:15.45] Exactly. Yes, Dan.

[00:40:19.53] DANIEL SCHWARTZ: So just going to what will be the impact on the practice of law, I do think-- everyone has to make their own judgment. But I do think there's a tremendous amount of change that is going to happen in the near term.

[00:40:34.60] And I think that it is impossible to predict exactly how that will play out because it is a byproduct of how the technology will change, how different people will use the technology, how different companies will change the technology and build on it. And also, frankly, laws and regulations, what will be allowed, what won't be allowed.

[00:40:54.81] And so, in my mind, the most important thing is for us, for everyone, to stay nimble and to think about both individually. How do you start using this technology? Where would where would you use this technology? How would you use it, becoming familiar with it?

[00:41:11.94] I think every lawyer, every law student, should be, at least, starting to familiarize themselves with this. I think that there is some amount of just time in building on that. And frankly, I think, organizationally, a lot of firms, a lot of schools need to be thinking about maintaining flexibility to be able to pivot. And I do think there are a lot of scenarios where maybe there are a lot of negative scenarios where maybe you need fewer attorneys. Maybe there are, are less hiring needs of big law, but then a lot of positive opportunities where maybe there's an opportunity to serve more people to more-- because you can more efficiently, whatever if you can write a will in an hour instead of 10 hours.

[00:41:57.24] Well, all of a sudden, now there are a lot more people who you can actually help. And so I think that there are huge opportunities. But there's going to be huge change and disruption.

[00:42:07.90] And I think folks need to start grappling with that now, both individually and organizationally. And if you wait too long, that might be a mistake.

[00:42:21.47] HARRY SURDEN: Those are some great points. And one issue that we hadn't talked about, but I think is important, is opportunities for access to justice. So Megan and I are working on a project at Stanford to help use some of these new technologies to help underserved communities who don't have access to lawyers to get help with some of their legal questions.

[00:42:43.02] So Dan, your question was great, that-- your answer is great that lawyers need to get involved. And you've recently written some scholarship about this.

[00:42:52.19] So what can lawyers do now? What should they-- how should they embrace these technologies?

[00:42:57.47] DANIEL SCHWARTZ: Well, that allows me to plug my paper. So thank you, Harry. So I have a few different papers looking at this. And I'm working on more.

[00:43:05.48] So one paper I have that is more just a way of, I guess, a first process for using this called "AI Tools for Lawyers, a Practical Guide" that's on Google. But essentially, it just walks people through some of the basic things, like chain of reasoning logic that Harry mentioned. But then it also sort of talks through how can you-- even really practical stuff.

[00:43:29.75] If you have a case that's too long to plug into it, how can you plug in that case into GPT so that you can actually get it to think through the entire case and analyze it? Or how can you ask it to cite the relevant provisions, something I alluded to earlier? How can you get it to not only draft the contract but then identify the ambiguities and then clarify the ambiguities?

[00:43:50.52] So it just walks through some basic techniques. And it's really designed as a way to get lawyers and law students to start actually familiarizing themselves with this technology. I think in a year. It will probably be very outdated if it's not already outdated now, which I hope not.

[00:44:07.70] But I think that the first thing to do is use this to-- frankly, spend the \$20 a month to get GPT-4 because it is really different than GPT-3.5. So I think everyone should be sending OpenAI their \$40 a month.

[00:44:24.08] HARRY SURDEN: Not Casetext, the 500 a month.

[00:44:25.52] DANIEL SCHWARTZ: Oh, yeah, I know. And Casetext if you can get it, if you can on their weight, pass their weight--

[00:44:29.27] HARRY SURDEN: We're also not sponsored by OpenAI.

[00:44:30.75] [LAUGHTER]

[00:44:32.93] JASON ADASKA: And start using this technology because there is a craft to it. And there's also just an understanding of where it will help and where it won't help and how can you use it better. So I think developing that skill set is an important first step.

[00:44:45.78] And then one of the things, obviously, I'm a law school professor, and actually, I came to this a lot because I started working as a fellow before I became a professor teaching legal research and writing. And so I think it's really important to start thinking about how to train our students to use this well. But I actually am of the opinion that the first thing we need to do is to teach them to do legal research and writing without this technology.

[00:45:09.77] I think it is-- one of the dangers here is over-reliance on this technology to the point where you can't understand what is doing well and what it's not doing well. And I've heard people make this analogy. And I quite like it.

[00:45:23.48] We teach kids how to do addition and subtraction, and multiplication before we give them access to a calculator. And I think that is very important because there's a way in which those skills are fundamental. And it's even important that even though we can trust a calculator. Well, imagine a calculator that makes mistakes some of the time.

[00:45:44.63] And so I think right now that pedagogically, we need to be teaching students how to do core legal research and writing, how to analogize and distinguish, how to write clearly, how to synthesize rules for multiple cases and apply those rules to [INAUDIBLE] facts in a sort of compelling fashion and leverage policy arguments. But then, once we have that foundation, then allowing them to use this technology to further refine that and make it more efficient. And so one-- going to one study that.

[00:46:18.06] So we did one study where we just looked at how chat those ChatGPT did on law school exams. And when we found it got about a C-plus. But it was able to get a C-plus in a variety of different areas where we just used a single prompt, single prompt that was used for all different exams. The paper is called "ChatGPT Goes to Law School." And it's already performing at the level of a not-very-good law student.

[00:46:42.95] But still, in an employee benefits class, in a constitutional law class, in a torts class-- so we have a new experiment that we're working on now, where what we're going to do is use GPT-4, but we're going to use it and have students use that in concert with their own skills and see what type of difference that makes in terms of their ability to perform. And the hypothesis, we'll have to see how things play out, is that GPT-4 is going to allow law students to not only perform much better on exams and analytical tasks, but we're also using-- we have a

separate experiment we're doing it on simple legal task, draft to contract, draft a memo, draft a complaint.

[00:47:25.31] And what we're I think what we're going to find, again hypothesis will see, is that this technology allows them to work much more quickly if they're trained well and to work much more efficiently. And so I think that it's a process. And we need to train law students. We need to train get lawyers to use this technology well. And that will take some thought. But I think that if we're thoughtful about it, it really does represent a huge change in how lawyers are going to work.

[00:48:02.03] HARRY SURDEN: That was a really thorough answer. And you made a bunch of great points, particularly about everyone should-- I agree should be at least trying out this technology and testing it, particularly GPT-4.

[00:48:14.27] And I will say there is a way to get a version of it for free using Bing Chat. It's not the same thing. But it kind of gives you a sense of what's going on. But don't put your private client data in there.

[00:48:29.60] Yes. Any other comments on that? What should lawyers be doing? Pablo?

[00:48:33.68] PABLO ARREDONDO: Yeah, again, I think steady as she goes. A lot of the stuff you guys are learning are doctrinal. And these are century-old principles and ways of legal reasoning that, frankly, aren't really impacted by technology.

[00:48:46.23] And I think you need to really have those down solidly. And I would say that I would rather have a mind that had to wrestle with the blank page from scratch, and clumsily futz around and a strikeout and then take longer, but then learn how to go from complete blank to an ordered system than somebody who thinks that they're, adding better adverbs to a draft that GPT-4 comes out.

[00:49:08.36] So make no mistake, you guys are not on the clock. No one's paying you per hour. Once you're practicing, there's other constraints. If I can do it faster, it's not about me.

[00:49:16.79] Your guys' number one job is to create the brains and minds that can advance the profession and serve the rule of law. And my personal Toobin on this is, yeah, learning how to, hey, have it make a draft, and I'll edit it. You know how to do that. It's just like if your friend gives you something to edit. It's suffering through that pain of that blank page. I don't--

[00:49:37.57] You might get around it. But ask yourself who suffers there. My two cents, probably not the view of Casetext's marketing team, actually, if I asked him about it.

[00:49:46.61] HARRY SURDEN: Yeah, that's a good point. It raises a larger issue that lawyers are not contract-drafting machines or

document-- they're advocates and problem-solving. Problem solvers help people through the legal system.

[00:49:59.52] So those skills, in conjunction with the basic skills, are still going to be necessary. Did you have a comment? No? So we have one more question, then we're going to open up to the audience for questions.

[00:50:15.32] So we hesitate to speculate, but we'll do it anyway. Where do you see this technology going in law or elsewhere in the next two to three years? And I'll just throw this out to the panel. And I picked two to three years because this is changing so rapidly. I don't even think five years from now we can do a reliable prediction, let alone 20.

[00:50:38.13] PABLO ARREDONDO: Yeah, I think it's going to be much more quickly adopted than anything we've ever seen. Law is a conservative group overall. But if what I've seen over the last seven months is any indication, they seem to be making an exception for computers that can read.

[00:50:49.86] And I think you'll see widespread adoption. I think that you'll start to see some fraying of the billable hours, some form of the business will change, and that probably will impact how many people are getting hired and for what. You might literally start to see that move pretty quickly.

[00:51:07.51] And I don't mean that in a doomsday way. I think there have be maybe different distributions of associates doing different things. And I think we'll all be use--

[00:51:16.77] Yeah, it'll be on our phones. It'll be just second nature for us to be using these LLMs to do the vast majority of things we're going to do. And I think we'll find it quite joyful. I think it's going to be a very wonderful feeling to have an AI that can schedule a damn calendar event, and then adjust for the Eastern time versus Pacific time, and do all of these things that if you actually add it up in our life, we're dealing with.

[00:51:41.43] Great. Dan?

[00:51:42.69] DANIEL SCHWARTZ: So I remember when I was a young associate in, I guess, 2004 doing discovery, and we didn't even have e-discovery then. And I remember sitting on my computer and literally just doing-- looking for keywords for hours on end, and then billing clients thousands of dollars. And I am like, how did I go to law school and do a clerkship to become a trained monkey?

[00:52:06.45] And I think that there's still a lot of that in the practice of law for young associates. Let's be real. And I think-- so I do think there's the real potential for this technology to allow lawyers and law students

to have more fun in doing their job and also to have more work-life balance.

[00:52:26.04] I don't think that this technology is going to put lawyers out of work. I just don't. Do I think it will change demand? Do I think that there may be some reduced hiring needs at some places? I do.

[00:52:43.68] But I think that there's actually a lot of ways in which things like soft skills are going to become more important. Can you communicate with people? Can you translate what's on the page into an explanation?

[00:52:57.60] Can you develop relationships? Can you be an advocate? Can you be a strategic thinker?

[00:53:03.63] I think those are actually going to be the more prized skills that lawyers are going to need to have, and law students are going to need to cultivate even in the next two or three years. And I do think the practice of law for many will get more fun because we can automate what is still a pretty grueling process in some element, like the discovery or produce a complaint that-- of the type that if you're just an auto accident lawyer.

[00:53:36.42] It's the same complaint. You're just copying and pasting funny things. You can do that now without having to spend an hour copying and pasting, or if you're writing something, and it's the summary judgment standard, you've written it 8,000 times, you can just tell GPT, OK, do that for me. So I think I think that there is a lot of hope but also a lot of risk, even just the next few years.

[00:53:59.11] HARRY SURDEN: So lawyers having 20-hour workweeks? We'll see if that happens. But no, those are some great comments. Jason?

[00:54:06.87] JASON ADASKA: Yeah. So I think the-- echo the statements of the other speakers, in terms of what the future's going to look like. I think it's definitely going to be everywhere.

[00:54:17.64] Right now, we see it in a couple of tools. I think it's going to-- in terms of technology for law, it's just literally going to be everywhere, either on the surface or underlying it.

[00:54:29.64] In terms of impact, I would expect that the transactional practices, especially those that have fixed fee models, are going to be most incentivized to figure out how to take advantage of that. So I think that's probably where they'll be in initial adoption. And one of the things that I think maybe seems strange to consider right now--

[00:54:54.46] GPT-4 came out in March. We all have a little sense of vertigo, of quickly, this has changed. I think we're going to be-- I think we're going to adapt. We're going to adapt and almost be bored of this in-- even by the fall.

[00:55:07.93] We're currently amazed by its ability to be able to address summary judgment or draft a patent or do analysis. I think people are very quickly going to mentally adapt to, OK, here's a set of tasks before that I had to grind through. Why am I doing that? I should be using some tools to either help me do quick summaries of things to do outlines, to draft small pieces. And I think it's going to quickly find its way into just things that people take for granted.

[00:55:41.02] HARRY SURDEN: Megan?

[00:55:42.75] MEGAN MA: I think, well, not only do I agree with the other panelists in terms of their speculation for the future, I think another interesting area of what large language models could do is its ability-- and Jason teased this, is its ability to tease out our implicit behaviors. I think a lot of our existing legal work is actually--

[00:56:04.44] We think about some of the best clauses that we've ever written. We tend to like to keep those to ourselves at times. And I think that if these models are integrated into our workspaces, such as our Google Workspace or our Office, we might start to see the habits in which we've taken over time and how we draft. And I think that that's going to be really interesting in terms of the future of being able to almost adapt and change the way that we act or behave as lawyers. And so I think that is going to be an area that is particularly interesting.

[00:56:39.42] HARRY SURDEN: Great there. Well, those are all great comments on our speculation. So let me open it up to the audience for questions.

[00:56:46.45] We have a tradition here at Silicon Flatirons that our first question goes to a student. So, Oh, we've got a student very eagerly volunteering. Terrific. And we encourage more student questions as well.

[00:57:00.27] AUDIENCE: Hi, my name is Christine. I'm a PhD student in computer science here at CU. I am wondering what you think some of the practical solutions might be to the current disparity in access to this technology, either directly or as a third party, fourth party beneficiary?

[00:57:18.68] PABLO ARREDONDO: We need more chips. We need better GPUs and more of them as fast as possible. I think that's, to my mind, one of the things that's making it very expensive, just literally running it.

[00:57:31.40] Maybe I wasn't sure. Are you looking for like a technical or a societal answer? Both?

[00:57:36.27] AUDIENCE: I heard whatever [INAUDIBLE] saying.

[00:57:38.36] HARRY SURDEN: Yeah, I have a thought, which is, there should be-- the government does need to get involved. This is a big enough technology like the internet.

[00:57:47.19] And I think we-- somebody said there's been ideas about having an Apollo program, and not just around the technology, around ethics, and governance. And there, I think there should be a public free option.

[00:58:03.36] This is an important enough technology down the road, like electricity or running water. Not immediately, but I think the government needs to get involved. Yeah.

[00:58:17.25] MEGAN MA: Sorry. I think also as well there's an important question around open source and licensing of certain underlying models. We've seen most recently, Dolly came out, which is an entirely open-source model.

[00:58:30.84] I think there needs to be more of that. And a lot of these models are built on LLaMA, which is a model that Meta put out there, but the licensing issues around that are not well-defined.

[00:58:40.33] And I think until we resolve those questions, then we can have definitely more space for free access to these models.

[00:58:50.73] DANIEL SCHWARTZ: I'm less convinced that it's imperative that everyone get their hands on these models and use them because I think that the value of these models really depend-- and the use really depends on context. And I think that one of the great things about these for lawyers is how much more efficient it can make law, and that then can actually have huge distributional consequences.

[00:59:16.86] So there's this well-known fact in the law that we have too many lawyers on the one hand, we have hugely not enough on the other because there's so many people who don't actually have access to legal services who need them for wills, for divorces, for custody matters. And I think that-- I'm not sure that we're ever going to get to a point where people are going to be able to or at least in the next two or three years, let me put it away, where you don't even need to hire a lawyer. It's just, Oh, I have a will, and I need this, and boom.

[00:59:47.19] There's some tools that purport to do that. But where I do think this can be transformative from a distributional perspective is maybe I only need to pay a lawyer \$100 to produce a will because, for them, it's literally just a matter of getting a few data points, plugging in and then just checking. And making sure that it's doing what you want. Do you think we need the human in the loop there who has some expertise? But that expertise can just be a matter of let me spend 15 minutes reviewing the output and fine-tuning it, saying, Oh, expand this, or add this, whatever else.

[01:00:23.59] So I think that distributional concerns may actually not lead to the, we need to give everyone access to this all at once as opposed to we need to make sure that folks are using these tools in a way that allows them to achieve efficiencies that can serve a wide subset of the population.

[01:00:50.49] JASON ADASKA: Yeah, so I'll just add. It's clear that this technology is a democratizing force for lots of specific information that previously people would have had a hard time getting to. The risk is hallucinations and whatnot, as people may not be able to trust all the information that's coming from it.

[01:01:10.38] So I think the bottleneck right now is anybody can go to Bing. And you can talk to it. You may or may not get what you want out of it.

[01:01:19.11] I think figuring out how to get the correctly curated versions of this out to the right people is honestly an open question. I think it's going to require the right experts and the right regulations in place to be able to get a vetted version of this for the different domains where it might be applicable.

[01:01:37.82] HARRY SURDEN: Another question from the audience over here.

[01:01:46.45] AUDIENCE: Thanks. Am I supposed to introduce myself? I'm not going to do that.

[01:01:53.00] So you've talked about how this is going to impact the practice of law from the, I would say, internal perspective. And I know this is like a whole other panel. So I'll try to ask the question, then give a more specific example.

[01:02:04.79] But do you have any thoughts on the current impact it's having? I can speak for myself personally, of advising your clients, especially if you're in the technology sector or work in-house on the using of these tools. Or is your company developing a product that uses them, which they probably are, by the way?

[01:02:27.36] And I think the context I'll give you just a narrow that question a little bit is software. You mentioned code generation tools, of which, of course, there's Copilot everyone's heard of, but there's many and models around that. I think that's most relevant to this audience, both from a legal and the technology sector on just on-- that's part of the practice of law.

[01:02:47.84] It is not just how we might use them internally, but how are we dealing with advising on both sides of that coin? So I just wonder if you guys had thoughts on that.

[01:02:59.23] DANIEL SCHWARTZ: I think the most obvious thing is you need to have a policy. You need to have a policy. So there are so

many employers out there that I think don't even have a policy for this for their employees, for instance-- or that just ban it.

[01:03:12.21] And I think, actually, in the very short term banning its use may not be a bad idea in certain contexts. It's not clear to me that-- given there's so much uncertainty, it's not clear to me that we want certain employees to be using it to do their job right now because we don't know exactly how to vet it. But I think that, at the very least, you need to address it.

[01:03:36.43] And I think that there are a lot of-- you need to do it quickly. But a lot of times, systems, over time, have developed to produce internal policies or internal--

[01:03:49.98] And they take time. And I think we don't have that time right now. You need to have at least a basic policy in place, like for universities or even for my students.

[01:04:00.15] For right now, I'm going to ban them using GPT on their exams, for now. And then maybe we'll have a class where we teach them how to use it and we're using it, but I think I think just addressing it is the first step.

[01:04:16.13] PABLO ARREDONDO: And I would just say educating yourself about, again if all you saw was people doing donuts on the lawn and going 90 on the freeway and backwards, you say no cars. How about that? Seems like a very good idea in that world.

[01:04:26.51] And I think, really going and understanding how these things can be used responsibly, which means secure servers, where the data is not retained, where you're not feeding into the model, where you're using you're coupling into a search engine to go over hallucinations, where there's guardrails to check quotes, all of these different things, I just think you need to learn about them because it's a very different world when you're using them correctly.

[01:04:49.17] HARRY SURDEN: Great. Another question from the audience.

[01:05:00.27] AUDIENCE: Thanks. My application is healthcare. I'm thinking about the very direct utility of your lessons to a physician trying to do the same kind of stuff with the same kind of problems. One of the threats that we deal with in medicine, though, is a direct-to-consumer application with, the bypassing the provider entirely. And they got access to these amazing tools.

[01:05:21.27] 100 years ago, we would say it was illegal for a consumer to have a stethoscope. That's ridiculous. Now you've got to can get an ultrasound. You do your own.

[01:05:33.93] It's not just democratization, but there's going to have to be some thought about. Do we have a way of making these apps safe

for consumer direct use, where the practitioner only learns about them late? You come into the ER. You're appendectomy is halfway done. And we need to figure out--

[01:05:54.66] [LAUGHTER]

[01:05:56.54] --where did Chat-- you know, leave off? So then, I would throw this back in because I don't know if you can answer that for medicine. Good luck to try. But in law, you got this pro se thing.

[01:06:09.82] And what happens when a murderer shows up in court and says yeah, I got my defense all prepared. Don't worry, judge. Here, I got it all here.

[01:06:19.38] I'll hand it to you. Ready. I'm innocent.

[01:06:21.30] So in any profession you'd like-- engineering, aerospace, [? corrections-- ?] pick the one the fun ones. What happens, though? We can't control this entirely as professionals? It's already long out of the--

[01:06:37.29] DANIEL SCHWARTZ: So let me-- yeah, go right up.

[01:06:38.45] [INTERPOSING VOICES]

[01:06:38.49] PABLO ARREDONDO: --let pro se people use our service. You have to be an attorney because of that. But also ask yourself, you can get pretty informed by a search result on Google and think you're ready to go into your search. Do you know what I mean?

[01:06:48.78] So ask yourself, how much of this is just consumers having access to information that they could then foolishly think suffices to make an informed professional decision? And how much of it is actually about AI? I think maybe it's a matter of degree.

[01:07:02.37] But we don't want people using it unless they're attorneys because I think it does give the illusion of maybe being more concrete legal advice than it actually is. It needs our attorney's oversight.

[01:07:13.95] AUDIENCE: So you're going to make it illegal for--

[01:07:15.54] [INTERPOSING VOICES]

[01:07:18.18] PABLO ARREDONDO: --to get our revenue up. We'll finally be--

[01:07:19.47] [INTERPOSING VOICES]

[01:07:21.27] HARRY SURDEN: Anyone else on the panel want to comment on that?

[01:07:25.41] DANIEL SCHWARTZ: I think licensor issues are really tricky. I tend to think that they've been abused In many settings to

actually protect incumbents. But at the same time, I think that they're necessary in a variety of settings as well.

[01:07:41.29] And so I think that we will continue to rely on licensure to ensure that, pro se is tough, but to ensure that, you can't just hire your friend to represent you and that you have to if you're going to have to get medical treatment, you have someone who's licensed and knows what they're doing. But as I said, I think it's tough. I think that we'll also see the abuse of licensure rules to protect industries that maybe should be shrinking.

[01:08:09.18] And whether that's law or not, I guess-- I'm not sure. But I do think that that's a possibility. And so they're just tricky issues here.

[01:08:17.85] I don't know. I really don't know how to-- what the right answer is. I think it's going to be very context-dependent. But I think licensure is the biggest answer we can provide, as well as warnings.

[01:08:27.99] And we have the warnings there already. Some of them are-- Google's more aggressive in its warning about what Bard will do, saying, look, don't trust this at all. This is completely--

[01:08:37.95] We don't have that for ChatGPT or GPT-4. They're actually fewer warnings. And so I don't know. I don't honestly also know how effective those warnings are. It tends to be most warnings are not that effective.

[01:08:51.21] MEGAN MA: I just want to--

[01:08:52.02] HARRY SURDEN: Oh, go ahead.

[01:08:52.72] MEGAN MA: I just wanted to go back to your question on medicine. So a while back, there was a tool called Babylon Health, which was trying to triage and diagnose these medical symptoms. And purport to say we have 92% accuracy, the average experienced medical professional of 30 years is 85%.

[01:09:13.78] But what we really get from this information is what do we really want from our professionals? Is it that when you receive this information, is it's accuracy the only thing that we're weighing? Doctors, we want that empathetic angle? What if you receive bad medical news? This machine is not going to give you that same empathy.

[01:09:30.97] And so I think when it comes to making that analogy with lawyers and whatnot, we need to really be rethinking what our role is as a lawyer that extends beyond the information that we are communicating.

[01:09:42.99] HARRY SURDEN: That's a great point. And one additional point I'll layer on top of that is we always want to weigh the benefits and the harm. So those are some real harms. But also today,

people are being harmed by not getting medical or legal advice that they can maybe get in this new world.

[01:09:59.83] Well, we are out of time. So please join me in thanking this terrific panel.

[01:10:03.90] [APPLAUSE]

Panel: Generative Art, Music & Intellectual Property Law

<https://youtu.be/OLckB23lmtU>

[00:00:00.27] BLAKE REID: All right, welcome back. It is lovely to see everyone today. My name is Blake Reid. I'm on the faculty here at the law school and one of the faculty directors at the center, and it's my pleasure to take us into a second panel.

[00:00:17.10] So we started practical today. We started talking about the role of generative AI and the practice of law. But now we're going to get weird, and we're going to talk about art and music and culture and intellectual property. And we have just an excellent panel to help us break this down.

[00:00:37.50] Immediately to my left is Casey Fiesler, who is an associate professor in information science here at CU. To her left, is Daniel Acuña, who's an associate professor in the Department of Computer Science here at CU. Daniel, welcome. And then our colleague from down the road at DU, Viva Moffat, who is a professor of law at the University of Denver.

[00:01:02.98] So Casey, we are in a moment. I wish I could say we timed this conference intentionally to be right around when the internet was just going to be losing its mind about the role of generative AI and art. So much is happening in the last week or two. We have the amazing Pope coat picture. We have the fake Drake song. We have fan fiction.

[00:01:28.74] The world is melting down. Situate us if you can in the cultural milieu that we're in and maybe if you could start leading us down the road of, why are we filtering this through the lens of intellectual property?

[00:01:44.97] CASEY FIESLER: Well, my first snarky answer to that question is because we don't have enough relevant laws about anything else.

[00:01:50.97] [CHUCKLING]

[00:01:52.80] I mean, besides, I think some other things that are relevant are going to be defamation privacy, not that we have enough of that. But we don't have a lot of laws around things like algorithmic

bias, for example. I also think that issues of ownership are just so-- immediately, it's something that people latch on to.

[00:02:14.40] So now that I think many more people in the general public have a basic understanding of how these systems work, even if it's just, oh, it's trained on the internet, which I think is a base level of what a lot of people realize here, they're like, I'm on the internet. Blake mentioned this. I'm on the legal committee for the Organization for Transformative Works, which one might expect fan creators to be very, rah-rah, fair use remix-- this is great.

[00:02:43.62] But actually, there's a lot of people who are very upset about their creative work being used to train these systems. So there's the input issue when it comes to copyright. And then there's also the output issue when it comes to copyright, which is around things like, how similar is this to other works, or who owns the copyright? And I also feel like a lot of these-- there are going to be legal issues, and there's already a bunch of lawsuits.

[00:03:15.18] But these are also like deep ethical issues, like, people stuff being used without their consent to create technology that then might replace their jobs. So I think that that's one of the reasons why copyright is coming up so much, is because it's the legal thing that seems to capture the ethical intuitions that people are having.

[00:03:41.47] BLAKE REID: So that's a perfect place for us to start. And before we dive down the rabbit hole on doctrine, I want to turn to Daniel. Daniel, with this notion that we might have intellectual property issues around the inputs and outputs of these generative AI systems, can you help orient us and maybe return to Professor [INAUDIBLE] introduction on what the aspects of the technology are that are relevant. How should we understand how this technology works for the purpose of intellectual property law? What are judges and policymakers likely to get wrong about this?

[00:04:19.11] DANIEL ACUNA: OK, well, thank you so much for the invitation. And I'm very happy to be here in such a timely conference. I'm not a lawyer. That's not my background. I'm from computer science, and I've been teaching AI, Machine Learning for a long time. And well, a couple of semesters ago, I taught deep learning. I applied deep learning. And one of the things that we discussed was precisely the technology that is behind GPT and ChatGPT.

[00:04:46.80] So I guess the basic idea of all these models-- so I guess I'm going to give a technical description of how these things work. Maybe that would allow us to understand the IP implications. So the basic way in which these methods work-- they try to find relationships between inputs and outputs. And those inputs could be images, and the output could be, what is in the image? What kind of dog is in the image, or something like that?

[00:05:11.97] The input could be audio, and you want to predict, OK, what is a text? So it's text speech recognition. Or it could be text. And the input could be text, and the output could be text as well. So you want the computer to find that relationship. So the inputs in this case for ChatGPT will be instructions, and the output is text. So you want the computer to learn the relationship between the instruction that you give the ChatGPT with the output that you want to generate.

[00:05:43.11] Now, these models fundamentally work on the idea that you want to predict the next word, although more precisely, the next token because you want to predict little pieces of text, subwords. So that task sounds simple. I mean, people have tried to do it for a long time. In fact, one of the first attempts was at the early 1900s when people were trying to build simple character predictions.

[00:06:10.78] So given this character, predict whether the next character is a vowel or a consonant. So that seems very simple, and you can get a lot out of that. So you can generate text that looks OK. And over the years, many, many decades, people created more and more sophisticated models.

[00:06:30.03] But we hit a roadblock, I would say, in the sense that, I guess, the complexity of language is such that we need very, very advanced methods to build this relationship between these instructions and the texts that we're going to generate. And over the last 10 years, we have these new models called neural networks that basically find these extremely sophisticated nonlinear relationships between the text, the previous text, and the next word that comes in.

[00:07:00.33] The issues that these models are so complicated that you need a lot of data to train them, and that means that we need data that we didn't have before. And we've been able to acquire them just because everything is more digitalized these days. So just to give you an idea of the data that we have-- well, we don't really know how ChatGPT works, but I mean, they released a paper, scientific paper. But it doesn't really explain much.

[00:07:29.82] But we know that it's using all of the internet. They are using the Common Crawl, which is basically like what Bing or Google would do to crawl all the internet and index it. We know that they're using scientific papers.

[00:07:43.71] Well, by using the internet, it's using all languages, all kinds of writing styles, and all kinds of qualities. And it's using scientific publications. It's using also code. So GitHub, which is this repository where people publish their code or open-access codes. They upload their code there, and we know that they're using that for also code generation.

[00:08:11.14] So I think one of the issues is that with these models, they are learning to make the prediction of the next token. And you get a lot

out of that. I mean, it's amazing to me that you can generate such sophisticated texts out of this, what seems like a simple task, just generate the next token. But these models are just so complicated and so complex that they're able to generate amazingly high quality text.

[00:08:47.15] But sometimes, they generate things that we don't know if they are true or not, so we have the hallucinations. And sometimes, they learn from biases that the text might have, so they might have not just biases about things that we care about, like gender and race. But they might have other kinds of biases as well.

[00:09:07.25] And in my research, one of the things that I study is potential plagiarism in tax and methods to automatically detect them. And one of the things that I'm worried about is that in ChatGPT and these generative models, you might have something that is generated. And it's really hard to tell or impossible to tell if it was taken from the input, so this huge crawl of the internet, of code.

[00:09:37.37] And even they have something called free law, which is a huge repository of opinions from the law. So we don't really know what's generating. It's a copy from the input. And we don't have methods, so we have a study where we show that detecting plagiarism with computational methods is really, really hard.

[00:09:54.39] So probably, everybody knows Turnitin, which is a company that detects plagiarism. But those are really basic kinds of plagiarism, so they're really hard. They're really bad at detecting paraphrasing, for example. So we don't really know if ChatGPT is generating paraphrasing.

[00:10:16.40] So I will just mention-- I don't know probably somebody already mentioned this. But for example, the creators of-- the people that are uploading code to GitHub are complaining that many of the code that ChatGPT and OpenAI Codex, which is this plugin that you use for programming, is generating code that is copyrighted. So it's taking pieces from the code that they never really wanted to be used by somebody else in this form.

[00:10:45.44] They argue that it's just using their code without their consent. So there are two class action lawsuits about this, and we have other cases outside of language. I know that we are in a law school, and language is one of the most important things. But we have the same thing happen in image.

[00:11:06.47] I think the title-- yeah, we have Midjourney. We used to have Midjourney, and I don't know. I miss it. But we have images, audio, and other kinds of media, where the same issues are showing up.

[00:11:18.20] BLAKE REID: Well, so Viva, I want to turn to you now. We've got this ecosystem of actors that's quite complicated. So we've

got folks who might have intellectual property rights on the data that's being used to train the model. We've got users of these tools. We've got the platforms that are administering these tools, then we've got the outputs that they generate.

[00:11:42.68] One might think there are some intellectual property issues that we can identify in there. Could you, in a few minutes, try and map out where do we start issue spotting? What's going on here? What disputes are happening now? And then we'll go from there.

[00:12:00.63] VIVA MOFFAT: Right, well, thank you so much, and thanks for having me here today. I think, first, I want to say that this is really, I think, useful to follow from Daniel's point, which is I think there's a lot we don't know and huge amounts of uncertainty. And in fact, some of the things that we don't know, I think, raise some of the biggest problems for lots of the disputes, but certainly for thinking about intellectual property and copyright disputes.

[00:12:22.94] And in the most broad sense, I do want to say that this is - I think, when we talked a little bit before this panel, I said, well, everything that's old is new again in copyright law. Not 100%, but the history of copyright law is the history of new copying technologies that cause a whole bunch of people, incumbent industries, creators, to freak out. And I think we're in that space right now.

[00:12:51.80] So the very first one was the printing press. That's why we have copyright law in the first place. And more recently, we've seen freak outs about the copy machine, I mean, the photocopy machine-- Napster, digital technology. And I'm not saying that AI isn't-- it's maybe as bigger and more revolutionary.

[00:13:11.57] But one reason that we may be talking a lot about copyright law is that copyright law is the thing, the legal thing that has responded, especially to these ownership questions that Casey raised, most regularly in these kinds of situations. So it is, in fact, the most developed set of law that we-- Well, I shouldn't say that, because somebody will tell me I'm wrong about that.

[00:13:37.41] But we do have this set of law that is used to thinking about, what happens when we have a new kind of way to copy things? And we said about the internet that it was costless creation or digital technologies, I suppose. We're just in a new round of even more frictionless and costless creation.

[00:13:58.71] And I think maybe I'll try to be really brief about this, and we can come back and talk about it a little bit more. I mean, I think that the major legal issues should be divided into two categories. And one of them has to do with what these models-- how they are using the data.

[00:14:17.12] And this is where I have some questions. But are they literally scraping and copying people's works from the internet? So how exactly are the data being collected and then used by the language models? And there are a few lawsuits that have been filed already.

[00:14:39.53] And even reading those complaints, it seems clear that the people writing the complaints don't know how the technology works, which is a common problem for copyright lawyers, I think. So I think that the question of how that information is being used, and whether that use is copyright infringement, that's on the input side, what's coming into these models.

[00:15:05.81] So one of the main arguments is that the use of images and text and music in the training data is copyright infringement itself. And maybe I'll put that to one side and then talk about on the output side, and I think we should talk more about the input side. On the output side, one of the issues is, well, what's coming out of these generators, the generative AI? One issue is, it protectable?

[00:15:34.69] So there have been people who've gone to the Copyright Office with their images or text generated by an AI, an LLM. And the Copyright Office has said in its new guidance, clearly no. If there is no human creator and contribution, then it is not protected by copyright law.

[00:15:57.10] And this follows on not allowing-- the monkey selfie-- the monkey to be the copyright owner. So it's not limited just to AIs who don't get protection, but any nonhuman entity cannot get copyright protection. So that's one output question in terms of copyright.

[00:16:18.25] And then the other one is the extent to which the outputs themselves should be considered infringing, how many of them create copyright problems or other problems. So there have been other doctrines brought to bear or that people have thought about-- right of publicity, DMCA problems, all kinds of other aspects. But I would say the primary focus is on copyright law and whether those things would be infringing. So why don't I stop there.

[00:16:48.68] BLAKE REID: All right, there's a perfect roadmap for the rest of this conversation. And I want to come back to Casey on the first question, Viva, that you asked, which is, is there something new here? Or is this a pattern in copyright law of which AI is just the latest in a series of technologies? And love to invite Daniel on this one as well. Is there something profound and new about this, or is this just the latest in a series?

[00:17:22.63] CASEY FIESLER: So I think the thing that is very similar-- so I agree with Viva. I think that the freak out looks quite similar to-- the first thing that came to mind was like, the VCR, so like photocopiers,

VCRs, Napster. When I talk about this in class, I'm like, here's the series of things that made the music industry lose their minds. And this is that.

[00:17:46.96] I think the difference is that there also is the other side of it implicated, where people are able to use this technology to create things that could be copyrighted, is I think maybe what's a little bit different because you weren't doing that with a VCR. So there's this side of it of people who are concerned about copyright infringement, but there's also people who are concerned about, can I protect this work that I'm creating? Could the work that I'm creating be an infringement?

[00:18:23.56] And I actually think there's a lot of ethical intuitions tied up in that, too, because people have very strong feelings about what should constitute creativity and what should constitute authorship. So I think that's the difference.

[00:18:38.41] BLAKE REID: Danielle, your thoughts.

[00:18:39.25] DANIEL ACUNA: Yeah. So I guess answering a little bit to what Viva was saying, I mean, the simplest kind of method will be just a statistical method, where you have, OK, given these words, what should be the next word? So you basically count. You go through the books and to the internet, and you say, OK, given that I have this word "the," should be the next word? And probably the word is going to be cat or house and things like that, so you have a distribution of possibilities.

[00:19:06.89] But the thing is that these models that are now introducing [INAUDIBLE], they have so much-- they're just so complex, and they have, in a sense, so much memory of what are all the possibilities that would happen after a text that they could, in principal, learn, memorize what you use for training. So even though it's still probabilistic, still you have a distribution of words.

[00:19:30.16] You could think of ways of-- force it to predict the most likely sequence of words, and that basically could account to just copying the training. We don't have a lot of evidence of that, so people have taken texts that ChatGPT generates and put them in something like Turnitin. And Turnitin and says, OK, this is not copied from anywhere.

[00:19:53.77] But again, it might be paraphrasing, so we don't really exactly know if that's happening. Actually, if you use things like image generation, which we don't have here, something like Midjourney or Stable Diffusion, which you say, OK, generate the image of a cat. It will generate an image, very high quality. And there are studies that show that it's actually copying the context, at least, of the training data set.

[00:20:19.48] So I guess, going back to what you were saying, Blake, about how is this different from before, so we have methods that are actually called generative learning. So you learn the distribution of

data. And these things have been around for a long time, since the 1900s. It's just that they were not that great. The kinds of things that you would generate were mostly data points, just two-dimensional distribution of data.

[00:20:49.66] But generating an image, for example, or generating text is something much more complicated. So I guess we have an amazing qualitative difference in what we are able to generate now. So that's, I think, why people are like, oh, my god, I didn't think this was possible.

[00:21:04.01] So it's almost like a percolation process, which we have this small, incremental progress in the kinds of things that we can generate. But because we just have this cumulative knowledge of how to have bigger data sets, better ways of training them, that we have this huge jump in the perception of the quality that we can do.

[00:21:23.87] But in a sense, they are the same thing. We just have a capacity to input much more complex data sets and output much more high-dimensional data sets. So I think that's what's making us be very amazed at these technologies.

[00:21:40.79] BLAKE REID: All right, so we have at least the perception of something new here. Viva, did you want to chime in on this one before we move on to doctrine?

[00:21:47.14] VIVA MOFFAT: Well-- I mean, I guess, I don't think I'm well-positioned to talk about, technologically, whether this is something-- whether it's qualitatively or quantitatively or both different. But I do think that this question of when a new technology comes up and it both allows potentially copyright infringement on a massive scale, which I think we have to say this does enable that, and it involves, it entails the ability to create new creativity or new-- I don't know, we should call it creation-- but new generation of work incredibly easily is we do have analogs for each of those in the past.

[00:22:34.31] And maybe this is a quantum leap forward or more. But I do think there are some references that we can think about in the past to help think through some of the problems.

[00:22:45.35] BLAKE REID: And for that, just to plug our technology journal, there's a fantastic article by Mark Lemley about a decade ago called "Is the Sky Falling on the Content Industries?" To which we could probably plug to ChatGPT, and say, and add and generative AI to the end of every paragraph. And it might describe what's happening here.

[00:23:05.00] But Viva, you've taken us to the question about doctrine, so let's go there. I think I just heard you say, copyright infringement at mass scale. So maybe let's start on the question of training the model and the inputs that are put into the model. And Daniel also added the users are providing some inputs. They're asking questions, so we've

got some inputs to deal with. Is that copyright infringement? And how do we evaluate that question?

[00:23:38.99] VIVA MOFFAT: Right, well, of course, the right answer, is it depends. That's the law professor answer, at least. And so I guess I want to be really clear about the thing that I'm thinking about and what-- there have been a couple of lawsuits, at least, already that I've looked at. And those ones are alleging that, for example, somebody's image or the image they've created or their text or their music is on the web or otherwise available.

[00:24:08.00] And this is my understanding of what the allegations are and that those things have been copied. I believe that they have been copied-- which is very important for copyright infringement, for the doctrine-- have been copied and used as part of the data training set. And you can tell me if I'm wrong about that.

[00:24:28.37] And the copyright argument is that that copying and use by either the-- I understand there are some entities that gather, actually gather those data, and that's separate from, say, OpenAI, that they just buy or use the data set from somebody else who has gathered that data set-- but that the copying and use of those materials for the training purposes of the AI is copyright infringement.

[00:25:00.18] So that's my understanding. That's the one big set of claims, is just in the training purposes, so nothing about the output of the AI. And now you want me to answer the question of whether that is copyright infringement?

[00:25:16.94] BLAKE REID: Before we go there, the ultimate question let me--

[00:25:19.46] [INTERPOSING VOICES]

[00:25:19.88] --Casey. Casey, any thoughts on this question of prima facie copyright infringement in training?

[00:25:27.76] CASEY FIESLER: I mean, I am a very rah-rah fair use person. [CHUCKLES]

[00:25:34.72] BLAKE REID: All right, we're going to get to rah-rah fair use in a second.

[00:25:36.36] CASEY FIESLER: Yeah. So I mean, I think maybe there's some question about whether a training model is fixing something in a tangible form of expression. Yeah. I probably prima-- probably. Yeah.

[00:25:57.94] [CHUCKLING]

[00:25:59.20] But I don't know that it's totally settled, and I agree that I feel like a lot of the people who are writing these lawsuits aren't describing how the technology works very well. But probably.

[CHUCKLES]

[00:26:11.92] VIVA MOFFAT: So I mean, the question I have is, are the models copying the works? Or are they just looking at them, like the way I might go and look on the internet? Are they just reading, which is technically I think not a copyright infringement? So I think that's a technical point that I don't know the answer to, but the people filing the lawsuits don't seem to know the answer to that either.

[00:26:37.84] [CHUCKLING]

[00:26:38.68] BLAKE REID: Daniel, do you have any clarity from the technical side of things?

[00:26:41.90] DANIEL ACUNA: They are not literally copying. It's not that they copy, and then they just produce that in the output. But they are using it for the training. So I don't know. It sounds very similar to what a search engine would do, just crawls the web, copies things. And then it tries to understand, when you search for something, if you like a web page, why you like that web page. So they try to find that relationship.

[00:27:04.72] But this feels a little bit different because, yes, they are they're using that for training data. But because these models are just so sophisticated and they might sometimes-- because the question that you're asking or a specific next word that you're asking it to predict-- it's just so specific that you might find, OK, I have this training data point that I use.

[00:27:25.54] So maybe I'm just going to follow that because that's the most likely thing that I should generate, this specific article that is just very, very, very, very narrow. And that's the most likely next thing that I should predict. So I'm not sure. I don't know the answer. But yeah. So it's complicated.

[00:27:45.40] [LAUGHTER]

[00:27:46.10] BLAKE REID: All right.

[00:27:46.87] CASEY FIESLER: Blake, I mean, we could have the same conversation about just scraping in general. Is all scraping that researchers are doing and common crawl, is that copyright infringement? [CHUCKLES]

[00:28:00.77] DANIEL ACUNA: So let me-- I guess, because I'm computer science, and I write codes, so sometimes you write code, and you put a license that is very open. You say, basically you can do whatever you want with the code that I write. But people are still suing GitHub, so basically the host of this code that is supposed to have a very open license.

[00:28:22.69] So people have a qualitative-- basically, it seems that they were given permission to use the code for others, for other humans to use it and reuse it in their system. But it seems that they are

not so happy with a company using their code to sell a product, such as Codex, so that other programmers can code. So I don't know if that's what's the argument there, but I think there is something that is perceived to be different.

[00:28:50.29] BLAKE REID: All right, so we've created a set of messes here, but we've only got a few more minutes to solve them. So we've got these questions about threshold, questions about copyright infringement. We've got in the GitHub case the stripping of content management, information, the question about licenses of content and how that all factors in. So we're not going to solve that, because we've got to move on to fair use.

[00:29:14.90] So Casey, you teed up the notion of strong feelings about fair use. Viva, I wonder if I could turn to you, though, to start us. Let's say that we have got prima facie copyright infringement here and probably at least an argument that that's the case. Is what these models are doing and what their users are doing with them-- is that fair use?

[00:29:40.16] [CHUCKLING]

[00:29:41.38] VIVA MOFFAT: That's all? I'll do that in just a minute or two. So I think, again, thinking about what the models are doing to the extent I understand it that is the-- can I use the word "scraping?" That pulling that in.

[00:29:52.39] If we think that maybe there's a decent argument that is copyright infringement, that is copying and use, then the question is, would it nonetheless be fair use? So you don't ask the fair use question until there is-- once there's copyright infringement, fair use is a defense to a claim of copyright infringement. And I think there's probably a pretty strong argument that the use of these models, the use by these models of the data for training purposes-- you can make a pretty strong argument, I think, that it is fair use.

[00:30:24.68] So one relatively recent example is the Google Books project, where Google copied-- so that's a lot of what these things are doing-- copied a whole bunch of books, millions of books, and not for just then redistributing the books, but for the purpose of allowing search and enabling people to find these books. And look at only snippets was part of the resolution, but I think that the end result of that was the copying by Google of those works was fair use.

[00:30:59.32] And that seems actually a fairly close analogy here, not perfect, but a fairly close analogy with what-- on the trains. I'm not talking about what people are creating with the AI, but the use by these companies of the copyrighted material and lots of noncopyrighted material for the training purposes.

[00:31:19.39] I think the argument would be, well, this is for a completely different purpose. We're copying these materials not for an expressive purpose, but for this purpose of training this new technical tool to do these other things. So I think, on that side, there's a pretty strong argument for fair use, not slam dunk.

[00:31:39.78] BLAKE REID: Viva, before we move on, so you restricted that analysis to the training part of it. And I we probably ought to open this up to the outputs as well. When we start considering what people are using these models to generate, does that change the fair use calculus at all?

[00:31:57.24] VIVA MOFFAT: Well, so then I think you have to do a new copyright infringement analysis before you do a fair use analysis. So somebody puts into Midjourney a text prompt to create an image of that. So the question is, is that action somehow copyright infringement? That's, I think, a really hard question. I think that's a really hard question.

[00:32:27.87] Are you creating-- I guess if you said, make me a version of-- let's assume the Mona Lisa is protected by copyright. Make me a version of the Mona Lisa in the style of Jackson Pollock. Is that then a infringing derivative work? So I think you have to get to the copyright infringement question first before you can get to the fair use question.

[00:32:52.38] BLAKE REID: And what's your answer on the--

[00:32:54.15] VIVA MOFFAT: On which one?

[00:32:54.75] BLAKE REID: Oh, no, on the first part of it, on the copyright infringement part.

[00:32:59.80] VIVA MOFFAT: I think that's really tricky.

[00:33:01.80] [CHUCKLING]

[00:33:02.70] That one, I think, is a hard-- so you're using this tool. It's not clear that there's any copying going on in the way that we ordinarily understand copying, when, again, this is-- I don't how it works, the technology, to some extent. But maybe all of that is a derivative work so that even if there's no copying, I'm creating a derivative work of-- I'm sorry, I'm not coming up with a good example of an image that is currently protected by copyright. [CHUCKLING]

[00:33:29.97] BLAKE REID: I mean, and just a really tricky example is the Fake Drake song this week, which is a new song that is performed in the style of Drake and The Weeknd, but is not a copy of any existing song. And so--

[00:33:45.36] VIVA MOFFAT: So then I think copyright does nothing for Drake.

[00:33:48.93] BLAKE REID: All right, so we've got these threshold issues of whether there's copyright infringement, and then we've got the fair use questions. All right, so Casey, you said, rah-rah fair use.

[00:33:59.56] CASEY FIESLER: I said that's how I usually feel.

[00:34:01.71] BLAKE REID: But it's almost as though you might be going in a different direction.

[00:34:04.60] CASEY FIESLER: Well no, I mean, I don't think that there's going to be any blanket answer for any-- I mean, like any fair use case, it's going to depend. But I do think this point about, you can't just think about the training process-- what you do with it matters for a fair use analysis.

[00:34:28.47] Google Books-- you're taking books, and you're creating a search engine, kind of like image search. You're taking the perfect 10 cover, and then you're not creating another perfect 10 cover. You're creating a search engine. But here you're taking a bunch of Drake songs, and you're creating a Drake song.

[00:34:47.22] [CHUCKLING]

[00:34:47.88] So I do think that that's a substantive difference. That said, I mean, if you had to ask me about it as a whole, I feel pretty positive about a fair use analysis for similar kinds of reasons. But I don't think that's how it's going to happen in practice. It's not like the world is going to sue OpenAI for copyright infringement. It's going to be Drake and GitHub.

[00:35:12.66] BLAKE REID: Daniel, do you want to come in on this one?

[00:35:14.17] DANIEL ACUNA: Yeah. So this is where my law ignorance-- so maybe I'll ask Viva about this. But I guess I like what you said, Casey, about what is this technology we going to use for? Because ChatGPT-- it's amazing. It's as a non-- I don't speak English as my first language. And so when I use it with other languages, it helps me a lot.

[00:35:40.69] So I think there's a lot of potential for good use of this. I think people perhaps are a little upset. I don't know if they're charged for this technology to be used. That's worse. Maybe that goes into the reasoning for the fair use. I don't know.

[00:35:56.49] I'm just reading the data set that is used likely by OpenAI and similar technologies, and they have a scrape of the web. They have publications. They have codes. They have open opinions. They have patents.

[00:36:13.74] And I study science, and I would like to use, for example, this data set on science. And I will not like to charge anybody for using the technology that I'm developing. I would like to do good things with these kinds of large language models. And I will just end by saying that

there are tons of things coming next that are not going to be-- that are likely going to be open.

[00:36:35.49] So there are going to be people not likely charging for these. It should be making these models specialized for certain fields, just putting them out there, not developing an infrastructure like OpenAI and charging \$20 per month, but just leaving it out there so that people experiment. So I don't know, in that case, whether that's fair use of these inputs to generate those outputs. But it sounds very exciting and something that we probably shouldn't constrain too much.

[00:37:02.26] BLAKE REID: So I want to zoom us out for our last few minutes here and talk about the implications of this technology for intellectual property policy. We've been talking so far about how intellectual property law might apply to these things, but we also should think about where it's going to go. So Viva, maybe I'll start with you on this one. We've long thought of the purpose of the copyright system, at least in the United States, as being a utilitarian one that's designed to spur the creation of copyrighted works.

[00:37:32.20] And I heard a statistic this morning that there are something like hundreds of thousands are on the order of hundreds of thousands of new AI generated audio tracks being uploaded to platforms like SoundCloud and that kind of thing every day. That order of magnitude will probably go up one or two as this technology evolves. Suddenly, we might find ourselves awash in creative works generated by artificial intelligence. What does that mean, or what should that mean for the copyright system?

[00:38:03.79] VIVA MOFFAT: Well, I mean, that's, again, really, a totally unknowable. But one thing we have seen when I said, oh, the history of copyright is looking at these new copying technologies is, almost always, there's been huge disruption, so really big change. Casey referenced the VCR, where the movie studios freaked out about-- freak out is the only word you can use here-- about the VCR. It turned out, for quite a long time, to ultimately create a gigantic market for them.

[00:38:33.22] Now, ultimately, that changed also. Over time, new technologies and streaming has come along and completely upended industries and markets. And this is not to say this is without even focusing on individual creators. And I think it's, this does seem really like the scale here of how instantly you can just create apparently endless numbers of Kanye's generated voice singing any song you want is-- whether you find that valuable or not, I don't know.

[00:39:11.98] [LAUGHTER]

[00:39:14.50] I mean, the scale of what's being produced is stunning and a little bit difficult to imagine how we're going to manage it. And so up until now and continuing, anything is protected by copyright, as

long as it meets a very minimal hurdle of creativity and it's fixed in a tangible medium of expression. One thing I wonder is whether this is going to push that question of whether we should have some kind of higher hurdle for copyright protection.

[00:39:43.75] I mean, it's almost like the system is broken if we have 10 billion new creative works created every year, so that's not a great answer. I mean, people are going to lose their jobs. New jobs are going to be created, new fields are going to arise, and other technologies are going to die. I mean, it is huge.

[00:40:08.02] BLAKE REID: Casey.

[00:40:10.63] CASEY FIESLER: So I'll first mention that the Copyright Offices started having listening sessions around AI and something that was raised a lot at the one this week was the potential for collective licensing, which is interesting. But the thing that I want to raise that AI that hasn't been brought up yet is the point at which something does constitute human authorship, I think, is going to be this big deal.

[00:40:34.61] So for example, the comic book where the registration in the image was canceled because the Copyright Office found out that it was created by Midjourney. That's not human authorship. It's like the monkey selfie or elephants painting, or whatever.

[00:40:53.68] So for those of you who haven't played with Midjourney, it's on a Discord server. So you can see all the prompts that people are using to make-- if you're using the Discord server. There is a huge difference between typing "girl in a red dress" into Midjourney and getting an output. I totally agree. Zero human authorship, should not be copyrighted.

[00:41:13.12] But I've seen people write five-paragraph essays of a prompt, tweak it a hundred times and come up with-- and does that constitute human authorship. At which point is prompt engineering actual human authorship? I think that is going to be a huge question because I do not think that you can say that those two things are the same.

[00:41:40.73] BLAKE REID: So just to tie this back to the first panel, it's interesting the way in which prompt engineering has been used to encapsulate a lot of really interesting and important human work and to look at the amazing things we can get these models to do. And we gloss over the fact that there is a huge amount of human input to get the model to actually behave that way.

[00:42:05.62] So we see it in a professional context when we're talking about the practice of law. We see it in these creative contexts. Daniel, final word before we go to the audience for questions.

[00:42:14.42] DANIEL ACUNA: Yeah, I really like that the idea of prompt engineering. I think the same thing happened with ChatGPT. I

mean, I'm an expert now in developing these very complicated and sophisticated tweaks, asking questions to ChatGPT. And there are databases of clever ways of asking to ChatGPT. So I don't know. Maybe those things are copyrightable. So I don't know.

[00:42:36.34] So I wanted to focus on things that I think are going to change from the technology point of view. I think people are going to-- my brother is an artist, if you will. So he's amazed at these technologies, but also he's thinking about ways of preventing the algorithms for describing the open work that he puts on the internet so that used for training.

[00:43:01.32] So people are thinking about creating standard ways of saying, hey, don't use this. Beam that part of the image, like the metadata. Or on a website, having a special text file that says, these are the permissions for training-- you have that for search engines. So there is this file that you put at the root of the website called robots.txt, and you say, please don't index my web page. These things are too sensitive, or something like that.

[00:43:30.44] So something like that's going to appear. I think we need to improve methods to the text generative AI. I'm very worried, especially since I work in science. There are paper mills in some countries that, basically, if you pay them \$50, they're going to generate a paper for you with images. And they look very good because the capabilities of these systems are going to increase over time. So we need better ways of detecting generative AI content.

[00:44:04.83] And we are going to adapt. So I think it's going to make things-- yeah. If you write good text that is grammatically correct, maybe that's not going to be that impressive in the future. So we are going to adapt, and I hope that also the law about copyright take all of these things into account.

[00:44:26.06] BLAKE REID: Well, a perfect note to open it up to the audience. And as always, under the wiser rule, the first question will go to a student. And I see at least a couple of my students have made the mistake of sitting right in the front row here, so cold-calling is going to happen. Are you a student?

[00:44:42.97] SPEAKER 1: Yeah.

[00:44:43.44] BLAKE REID: All right.

[00:44:43.85] [CHUCKLING]

[00:44:44.48] First question to you.

[00:44:48.17] SPEAKER 1: Yeah. I'm wondering with the issues of authorship, an AI generated image is not have an official author. Then who is liable for the copyright infringement if a generated image is considered copyrightable?

[00:45:05.04] VIVA MOFFAT: I mean, so I've actually been thinking about this question. And I can't quite figure it out. If it's not protected by copyright, because there's no human authorship, but if somebody has generated it, is the person who generated it or put in the prompt the proper defendant?

[00:45:27.90] Is OpenAI, or whoever, the defendant? I don't know. I mean, I assume at this point the law would say, oh, the person who made the machine spit out that image would be the defendant for that being-- for the copyright infringement suit. And I think it would prompt some of these questions, and I teased out a little bit.

[00:45:50.07] Well, was there any copying that went on? Or is it a derivative work? Well, certainly could be if I said-- if the prompt was, oh, copy take this photograph in the style of X painter. So I bet that that's what the law would do, but I haven't seen that happen yet.

[00:46:12.33] BLAKE REID: So I want to add-- these sound like really technocratic from the perspective of copyright law kinds of questions. From a policy perspective, are these the right questions to be asking if we're trying to figure out who should bear the costs of infringement?

[00:46:30.21] VIVA MOFFAT: Yeah, I mean, you give a copyright lawyer a hammer, and everything looks like--

[00:46:33.63] [CHUCKLING]

[00:46:33.84] --copyright problem. So I mean, I think maybe lots of things that we've talked about here make it really seem like maybe copyright is not a capacious enough tool to address some of these problems. And some of the other claims that have been brought are akin to-- so in the Getty Images suit, they have their Getty Images watermark that does show up in a lot of the generated images, sometimes warped and weird and stuff.

[00:47:04.41] So they've brought trademark and unfair competition claims. I think there have been right of publicity claims that have been brought. So it isn't just copyright that people have been thinking about, but other kinds of-- did I say right of publicity? And then there are huge privacy concerns, of course.

[00:47:24.54] BLAKE REID: So Casey, so you started us off with this point about other areas of law. Can we return to that? , If not copyright what should we be thinking about instead?

[00:47:35.76] CASEY FIESLER: I mean, there certainly are proposals for laws around governing AI, more so in other countries than here. Or we have OSTPs blueprint for Bill of Rights for AI. It is possible that something like that could turn into-- I mean, do I think it will? No, but it's possible that something like that could turn into law.

[00:48:01.20] And it's very rights-based. What are our rights as humans to have transparency and algorithmic fairness and to not have our jobs replaced and these kinds of things? So that would be the obvious thing, is to actually govern AI, to try to regulate AI in this country, which-- maybe now.

[00:48:28.17] And then there are other weird things that are popping up. There's been at least one defamation lawsuit already, which I think then brings up the does section 230 apply to ChatGPT question? The answer is probably no. But that's coming up.

[00:48:44.79] And the question is, does the fact that OpenAI has this disclaimer that there might be wrong information? Does that make them not liable for that kind of thing?

[00:48:56.22] BLAKE REID: Excellent. All right, another question. I'll go to a student first if there are any student questions. Also, I will open it up. Is that a student question back there?

[00:49:06.47] SPEAKER 2: Hi I just want to preface that I'm not an IP student. I'm interested in criminal immigration, but this is very interesting to me. I'm interested in the music aspect because it seems like art is really subjective. But like I saw talk of Kanye West singing "Rolling in the Deep." And I was like, this is ridiculous. But you know--

[00:49:23.84] BLAKE REID: I think "Hey There Delilah" is the better one, but we can discuss afterwards.

[00:49:26.97] SPEAKER 2: I'll have to look for that one later. But that's their name and their image. And you know what happens when you have a really soft artist singing, like WAP by Cardi B. and Meghan Thee Stallion. What does that do to their image, and what kind of lawsuits do you anticipate being brought? Because everyone knows what their voice is and what the other song is, and so it seems very complicated.

[00:49:48.88] BLAKE REID: Viva, do you want to take this right of publicity question? Maybe to give you a chance to think. I mean, there's this intuition I think that-- and I highly recommend for folks that don't know what's going on here, go listen to the Vergecast episode this morning. And they do a demonstration of how to make basically a track using-- they do a track with Drake and with Eminem and with Jay-Z.

[00:50:14.45] And it takes just a minute. It really is pretty easy to do. And I think there's this sort of revulsion or amazement that when you see this happening, it's like, oh, my gosh, you can just take this person's persona and play creative God with it. What is the law have to say about that?

[00:50:34.70] VIVA MOFFAT: Well, the law mostly has said, you don't have a lot of rights in your persona or your name. So it's really a state law and state-to-state differences in rights of publicity that grown out of

rights to the right to privacy. So it is not a very robust kind of legal claim. But I do think we have some feelings of intuition about, oh, that seems weird. And if it were me, I might not want that happening.

[00:51:06.35] And echoing Casey's point about, are there some ethics to this that we feel that we care about, but there aren't the rights of publicity. I mean, occasionally sound-alike has been allowed. A sound-alike claim has been allowed, but those have not been very, very effective ways to police that. I mean, I actually have to say, I think it's an open question of whether we think you should be able to control that, your name.

[00:51:33.18] And I mean, the use of the sound of your voice is-- I think that's an open question about whether we're better off as a society, which is what copyright is trying to answer in general, is like, how much creation do we want? And how much do we want things tied up in rights?

[00:51:52.56] CASEY FIESLER: I also think that outside the right of publicity, the US does not have very strong moral rights as related to copyright in particular. So when you talk about market harm and fair use, for example, that doesn't mean this is going to embarrass you.

[00:52:09.77] So an author can't get upset about there being racy fan fiction written about their world. That is like a moral right saying, we don't have anything to protect that really. So the example of, oh, Harry Styles singing WAP or whatever-- it's not going to be a legal issue on the fact that, oh, that's embarrassing and will make him look bad, at least on a copyright perspective.

[00:52:39.89] BLAKE REID: Well, I just have to make an announcement. Previewing next year's programming, we're actually going to be ingesting all of the audio of these panels today. And we won't be here next year. It'll just be--

[00:52:50.12] CASEY FIESLER: Yeah, train the whole thing [INAUDIBLE].

[00:52:51.86] BLAKE REID: Yeah.

[00:52:52.28] CASEY FIESLER: I don't need me any more.

[00:52:53.09] BLAKE REID: Or YouTube your TikToks. Perfect. All right, I think we've got time for one more question from the audience. Yeah? It looks like perhaps another student question.

[00:53:03.67] SPEAKER 3: Hi, my name is Jacob. I'm a graduate student. So in the last panel, it was mentioned how good GPG4 already is at understanding laws and making judgments or assessments on them. And I know a lot of this still is up in the air, undecided. But in theory, if we drew the line and made some copyright laws here, you could give those laws to something like GPT and tell it, hey, generate

an image but keeping in mind, don't create something copyrighted. How would you reconcile that type of situation?

[00:53:39.16] BLAKE REID: We have robot creators versus robot lawyers.

[00:53:42.55] [LAUGHTER]

[00:53:43.66] Who wins? Go.

[00:53:48.97] CASEY FIESLER: That's an interesting example. So a lot of these systems already have safeguards in them. So for example, when I was playing with ChatGPT, I asked it to write me erotic Batman fan fiction. And it refused both on the grounds of content and on the grounds of intellectual property, which I thought was interesting. So in theory, these kinds of things could be built into the system if we knew what these answers were.

[00:54:17.71] DANIEL ACUNA: Yeah. I guess, yeah, that's a very interesting question. I think it's going to happen. I think there is some discussion that ChatGPT or GPT4 for reached the limit of how much data they can ingest and how big the model is. So there is some discussion about how much more we can extract with just scaling these models.

[00:54:37.78] Or we don't have more internet, unfortunately, so I don't know what more data we can use. So answering more complicated question will require some time to make them reason more with more sophisticated reasoning. But I think, in general, we are going to adapt. I mean, I see that all of these-- I mean, all of these laws are going to start becoming outdated.

[00:55:06.04] I mean, when Photoshop came out and we were able to modify things and modify the image of anybody and people got used to it, said, OK, this is Photoshop. I won't believe this. And I think something like that will happen if we see something written on some essay denying global warming or something like that and it's very well-written and lots of good citations, we're going to add some doubts, well, maybe this was generated or something like that, the same with flat Earth and things like that.

[00:55:37.18] And with elections, too-- so we have an election next year, and we're going to see probably lots of nonsense on the internet just facilitated by these technologies. So hopefully, we'll adapt, that the laws will adapt. And just, people will become used to this, unfortunately, to this noise.

[00:55:55.69] BLAKE REID: Viva, just quickly, before I go to you, there's an amazing picture out of one of the image generation of Donald Trump and Barack Obama playing basketball. When we start to think about how it might come up in the election, very, very interesting. Final thoughts.

[00:56:12.97] VIVA MOFFAT: Well, one response I had to that question was that I think we absolutely need to think about how we might be able to use these tools to help solve the problem. I mean, I think that's what you were getting at, is can we-- maybe they can actually help us, but I also think we need to be really, really careful. And maybe this really leads to the next panel.

[00:56:36.14] But we're careful about how much we want to give over to the tools, as I don't want a system that is going to tell me I can't make fair use or be something that just automatically decides that my use of something isn't fair use. And that's a very minor point, but I think that raises all kinds of much more significant concerns about how AI operates.

[00:57:01.72] BLAKE REID: All right, on that note, we are standing between you and lunch. So please join me in thanking our panel for the excellent discussion.

[00:57:08.40] [APPLAUSE]

Panel: Potential Positive/Negative Impacts of LLMs on Governance & Society

<https://youtu.be/hPvXmqJJ5LM>

[00:00:00.15] HARRY: Really important panel. And we have some amazing experts who I'll introduce in one second. But I just want to point out the fact that, I think, it's really important to look at both the positive and negatives of this technology and not just fall into the trap of only thinking there are positives like some people do or only thinking it's negatives. And also not catastrophizing, right?

[00:00:22.95] There is a well-known tendency of the brain-- the human brain, to go to the worst case scenario. So what we're going to try and do is a really balanced-looking at both the positives and negatives of this technology, which will be disruptive in both good and bad and unpredictable ways. So joining us is amazing panel.

[00:00:46.18] First, we have Professor Martha Palmer, who's a Professor of Computer Science and Professor of Distinguished for Linguistics at the University of Colorado in Boulder and has a rich history in working in exactly this area of natural language, processing, and understanding. Next, we have Eliana Colunga, who is an Associate Professor here at the University of Colorado Boulder, and also somebody who has spent a lot of time thinking about these issues.

[00:01:16.36] And last but, definitely, not least, is Orly Lobel, who's the Warren Distinguished Professor of Law and Director of the Center of Employment Labor Policy at the University of San Diego. And just came out, not that long ago, with an amazing book on AI and policy and

governance generally called the Equality Machine, which I highly recommend. So let me dive right in for our first question here. And I'll pose this to Orly, and then open it up to the rest of the panel.

[00:01:50.62] So Orly, what are some ways that you think of large language models in these new generative technologies like GPT-4 more broadly in terms of creating improvements for society?

[00:02:02.43] ORLY LOBEL: Yeah. Well, Thanks. First, that's a really big question. And what you said at the outset, I think, is really important to look at potential and risks at the same time. And even more importantly, to try to envision the potential. So just not take it as a given. So I think throughout the day, we've already heard about a lot of great uses.

[00:02:32.31] We talked, most specifically, about access to justice and looking at low-tech, thinking-- most of us, probably, in the room are attorneys thinking about how some human functions can certainly be replaced by generative AI where, certainly, the more mundane functions can be taken away and, perhaps, give us more time for creativity and more time in general.

[00:03:06.75] But also just thinking about all the people who can't pay the costs of attorneys, can't pay the cost of medicine. So there were other examples, I think, given throughout the day about health care radiologists bots that are performing surgery. There's so many different uses. And one of the things that I want to say at the outset that has been important for me in my research is that, when you asked me, what is the potential, what are the good uses, what I always want us to think about is a comparative advantage.

[00:03:45.27] I see a lot when we're talking about Chat GPT and its flaws or any other program that uses AI. I see a lot of this double standard where there's an expectation of perfection from the machine, rather than asking, is it an improvement, is it outperforming what we have now? So, I think, the most intuitive one, like example would be, a self-driving car, an autonomous vehicle, where you see the media reporting, oh, it got into an accident, so it's unsafe.

[00:04:23.88] And, of course, the question is, is it safer? And so similar in a lot of fields, specifically, the field that I work in the most, and there I think there's already so much use with AI is in hiring, there's a lot of talk about AI bias and what are the outcomes when you tell a software to do all kinds of resume parsing, applicant sorting, video interviewing, facial recognition, emotional recognitions. There are lots of technologies.

[00:04:59.35] And there's a fear that there's going to be disparate outcomes. And again, the right question always to ask is, is it better-- it can do better than the human biases that we already have? And I mean, we have a history of a lot of exclusions everywhere. But again,

specifically, I study the labor market and how hiring happens and how networks are closed. And very skewed in a lot of ways professionally.

[00:05:32.01] So I think the potential there is really to look for underserved talent, undervalued talent and to be more inclusive.

[00:05:42.25] HARRY: Yeah. That's a really important point. And I just want to emphasize the idea of always comparing it to what we have today. So it turns out that humans also make up facts, not just large language models, all right? And humans have biases, and if you have an error, an a legal document. Humans make errors, too. So those are some great comments. Let me throw it out to the rest of the panel on some of these positive benefits.

[00:06:11.32] ORLY LOBEL: You want other possible benefits?

[00:06:14.03] HARRY: Yeah. For now, then we'll look to the other side.

[00:06:20.98] ELIANA COLUNGA: Yeah. Go ahead.

[00:06:22.85] MARTHA PALMER: OK. So I, certainly, like all of the things that you suggested. One of the areas I've been working in more recently is education and trying to put an AI partner in a classroom with students. And the more I'm learning about GPT-4, the more excited I can get about the potential for how it could just transform learning experiences.

[00:06:50.93] So a kind of a vision that we've had in the field for at least 20 years or more now is being able to take a good textbook that is written for a particular student and take that material and make it either-- make the vocabulary either simpler and more accessible or possibly make it a bit more sophisticated for an older student or make the material have more images and less text or less images and more text.

[00:07:24.55] Or even potentially, translated into another language. And as far as I can tell, GPT-4 should be able to do all of that. Now, it's not going to be able to do it immediately. And we'd have to get some funding from somebody to work on it and figure out how to make that happen. It could easily turn the books into audio books, for instance.

[00:07:48.73] So there's a wonderful potential there to just take any particular type of informative document and just produce all these different versions of it. You could ask-- people could say, I want this same information, but I want it in Spanish or I want it in almost all visual and very little text. And that could really improve people's learning experiences.

[00:08:24.13] And you could have tutors. And again, we're not there at the moment. It would take a bit of tinkering. But you could have sort of personalized math tutors and chemistry tutors and foreign language tutors and just any subject you can think of that could be one-on-one

with a student and meet them right where they are, answer their questions, and give them new material and new challenges that are perfectly suited for the level that they have. So--

[00:08:57.01] HARRY: Orly, did you want to respond?

[00:08:58.14] ORLY LOBEL: Yeah. No. I love that example. Actually, in the book, in the equality machine, I look at tutors in schools. And again, this is an issue of access. I know that there's a lot of resistance I've heard. I've been in debates with the ACLU saying, oh, these companies are giving free laptops or free iPads to schools. They're just going to try to extract the data.

[00:09:26.17] But really, what we're seeing is, especially, when it's coming out of academia, people like you who are working on these projects, I looked at a project out of MIT that's looking at social robots. So these are fuzzy humanoid robots that have-- before GPT. But they have some degrees of machine learning in them and ability to personalize-- they use facial recognition and they do exactly what you're saying.

[00:09:59.75] But also they model fallibility and they model learning in ways that the teachers that are strapped for resources, they have so many students and so little time. They just can't do that. And so the way that you presented it as tutors, again, a lot of times, we think-- and I know that we're going to talk about this in the panel.

[00:10:25.22] Like, are we going to be replaced versus are we going to be aided in our work? And that's what I tried to suggest with attorneys as well. And we certainly can see this happening with teachers. I think that what we really need to think about is how our own work can be enhanced, can be helped rather than supported. And the goals achieved rather than just thinking about it as like replacing teachers.

[00:10:52.22] HARRY: Yeah. Great point. I think it's ability to really impact and improve education and the visions you both spelled out are going to be really important. You can also get it to explain a complicated academic paper to you like you're talking to an eighth grader. And it'll do a good job taking a paper that at least I couldn't read and then explaining it to me in really understandable terms. And I think this can democratize knowledge to people. It's going to revolutionize research, I think, in important ways. Eliana?

[00:11:27.86] ELIANA COLUNGA: Yeah. I think personalization is a really encouraging thing that it could do. I think it is very exciting. The possibilities there. Some of the stuff that we are working on in my lab is precisely on adapting things for tiny little toddlers. And so there's that thing about hitting the right level of explaining information to different groups of people.

[00:12:00.24] But there's also other types of personalization that you can imagine having, right? So maybe you want your audio book read by Kanye West, right? Or maybe you know a child who is really into one specific topic. Can learn about a concept within that specific topic that they find already interesting and engaging. So I think the possibilities are endless.

[00:12:27.84] I also think in terms of bias. One of the things that these models-- I mean, large language models, just like large models, do very well is take vast amounts of information and integrate them in a way that makes sense. And so you could see using this type of model as a way to catch bias, as a way to self-monitor but not just monitor the models, but also just monitor what is going on and pinpoint areas where bias might be interfering with the processes.

[00:13:10.76] HARRY: That's a really important point. I think, this technology can be used to detect latent biases that we haven't seen before and help improve in those areas. I really like that idea. Do you want to have a follow-up, Martha? Yeah. Go ahead.

[00:13:26.57] MARTHA PALMER: Can I offer a less profound potential benefit of GPT-4? This is one of my personal favorites, because I've just spent the last six months trying to shift from Comcast's TV and phone and internet to CenturyLink. And I am now a veteran of at least 30 hours of online chats and outsourced service centers.

[00:13:52.67] Imagine having an online chat that actually worked, that in five minutes would solve your problem. And that's totally something that these things could do.

[00:14:01.91] HARRY: That is a great point. That alone, could improve national happiness by 10%. OK. OK. So we've heard a lot of positive views, but we also want to look into the negative aspects. So let me start with Eliana. What are some of the risks that you see in terms of disruption in society of this powerful new technology?

[00:14:22.62] ELIANA COLUNGA: Yeah. So I want to say like straight out from the onset that I'm not going to talk about super intelligent AI deciding that the world is better off without humans and poisoning all of our water sources or anything like that. I don't think that as catastrophizing. I think that's going beyond, I think that's a silly idea. We can talk about that more if you want to.

[00:14:51.64] I think a lot of the risks and the-- yeah. The pitfalls of this technology come from misunderstandings about how computers work, misunderstandings about how AI works, and misunderstandings about how people work. And so there's a lot of things that are problematic right now given the way things are. But there's no reason why things have to stay that way.

[00:15:28.46] And so I want to highlight two potential issues. And I'm highlighting them, because these are already existing problems. And so I think they can be made worse with this technology. But I also think they can be helped. They can be ameliorated. Maybe corrected with this technology. And so I cheated in my assignment. I'm just an optimist. And so one of them has to do with the homogenization of language. And not just language, but language is just more than just words or just sentences, right?

[00:16:21.01] Language is stories, language is culture. We transmit our values and our culture through language. And we codify it in language. So language goes beyond that. And let me give you a concrete example of the thing that I'm talking about. So a few years ago, I was at a workshop organized by the HeadStarter Network. They work on innovation and technology in the early childhood development and education.

[00:17:08.02] And so they're affiliated with Head Start, They're not-- sorry. They're not affiliated with Head Start, but they work together, sometimes, and they have the same goal of positively impacting the lives of vulnerable young children. And so this workshop was on trying to develop principles for ethical use of AI in this particular domain.

[00:17:39.47] And so one of the use cases that we were discussing was, let's have AI just, basically, listening to everything children say and these are tiny, right? These are preschoolers. They're not going to be talking about anything that could get them in trouble. So it's not a privacy thing. But basically, we can use this to build normative models of what language and cognition should-- what are the stages that it-- how it should proceed.

[00:18:11.35] And then when a child deviates from that, we flag them for intervention. And that's really scary. That's really scary because there are different ways to get at the same place. There are a lot of variants of language. And there are different ways of speaking and developing that are all acceptable and should be embraced and celebrated.

[00:18:35.96] And so when you have the model as one standard way, as the ideal, then it becomes really problematic. This is a problem that already exists. And we already use standardized tests for this purpose. So it's not something that AI would introduce, but it's something that I could make worse. And I think, because of the way we misunderstood AI as being objective and perfect and things like that, I think it's potentially a problem.

[00:19:15.41] The second thing I want to highlight is the devaluing of human labor. And again, we already do this, right? So every time I see one of those videos of someone expertly harvesting produce or cooking, I just want to scream, unskilled labor what? This is not unskilled. You tried to do it. I know I couldn't. And so we already do

that. And I think we're all of excited about a future in which repetitive, boring manual labor is not around anymore.

[00:20:00.20] I know personally I don't want to pick up any more socks off my floor. That'd be great if we had a sock picker upper. But people start getting worried when it's like, doctors are getting replaced, lawyers are getting replaced, scientists are getting replaced. And so, I think, again, this is a problem that already exists.

[00:20:25.17] And I think we need to make a thoughtful conscious decision of how we want to spend our time. That is what matters here, right? How do we want to spend our time? And what do we want our tools-- the tools that we are designing and creating to help us spend the time the way we want to. Maybe it is making cat videos. That's fine. You'll find your bliss. But we can just be thoughtful about it.

[00:21:00.80] HARRY: Wow. Those are really two profound points just to summarize. Your first point was really interesting, because there might be this tendency to collapse, nuance, and context, and take all the shades of gray and have AI collapse it into one automated standard. And that's not necessarily a good thing. We heard this in the last panel. A version of, why can't I just decide fair use or something like that?

[00:21:27.78] So I think it's going to be-- we're going to be something we're on guard about that we don't-- in the race to efficiency and certainty, we lose some of the broad spectrum of humanity and human nature. So that was a really good point. I love that. The second point, I think, you said, which is also terrific, which is, we should have always been caring about people whose livelihoods are being automated away.

[00:21:53.30] As a society, not just when it starts coming after knowledge workers and things like that. And I think you're absolutely right about that. Let me throw the same question about risk out to the other panelists. Orly?

[00:22:09.99] ORLY LOBEL: Yeah. I'm happy to follow up on that, because I agree with Eliana's points. And it's actually something that I've been thinking about quite a bit. Exactly like you say, these are not new problems. Well before we had Chat GPT and before we had integration of AI and so many different aspects of our production, and before we had digital platforms. So there's a lot of this spotlighting on Uber as the gig economy replacing full-time employment and changing the nature of employment relations.

[00:22:55.47] But the move to contingent work has been something that has been happening for decades. And it did start with the lower skill, then moving up to every single job. What I have argued in my research for a long time is that, it's always been a political decision and

an anomaly and a historical path-dependent decision to tie so much of our social welfare to work and to a single employer.

[00:23:33.27] It has not been the case everywhere in the world, but it's very Americanized. And so it's spread to Europe and other places. But I think this is a real moment where we should question a lot of those assumptions. That our Social Security, our welfare, our health, our reskilling, and investment in human capital. Our retirement. That all of those are tied to-- and also just worker-- all these things are tied to what is our job right now defined as--

[00:24:17.58] And that's a completely political decision. And as a big piece of it, we've also very much politically, historically, and now still today, tax much more of labor than capital. And so we are at a moment where it's a perfect storm where these political arrangements that we had-- now, that they're getting more to the top, I think maybe it's a moment where we can question them all the way down and think about issues of global justice and distribution.

[00:24:58.15] So these are big answers. Like well, let's change the system. But what I want to say-- and Harry, you said that, in a nice way, it's like there's been all these cycles of automation. Like industrialization is also a cycle of automation. We're seeing it, maybe, at a different speed right now and on different scale and different kinds of jobs.

[00:25:22.93] But we've always had this. We talked about this in the copyright panel as well of there's always the sky is going to fall, there's not going to be this concept of work or creativity or-- and it turns out that that's not true. But it is true that we need to understand that it's not on technology to get it right that we have values in a democratic society that everybody has the right to human dignity and livelihood and to sustain their lives and their well being. And that's on the democratic process, not on the technology.

[00:26:04.19] HARRY: Absolutely. That's a great point. And then part of what, I think, is reflected in there is, in these technological trends, in the long term, often, society adapts and new types of jobs are created that we couldn't even imagine. Like nobody could have imagined the computer game designer 100 years ago, right? Whereas human calculators were obsolete by the computer.

[00:26:27.22] So in the long term, jobs don't go away. We figured it out. But in the short term, people, individuals are disrupted. They can't pay the bills, they can't feed themselves or their family, and we as a society need to be sensitive to that and do something about it. We can't just ignore that. And I think that's a great point. Martha, did you want to weigh in?

[00:26:46.98] MARTHA PALMER: Well, that's a hard act to follow, right? But I think there's one more point to make up. And we're talking, I

think, in general terms about the idea that GPT-4 could do away with of different kinds of jobs. And that's certainly true. I mean, there's been a lot of talk about how good the programming is.

[00:27:09.74] And our computer science department here, the people who teach the introductory programming class are in dire straits, because GPT-4 can get everything right on the final exam for the programming class. And how do they test the students in such a way that they can make sure they really know what they're talking about?

[00:27:32.52] And that's also going to be true in a lot of other areas too. Like editors, editorial assistants, GPT-4 can do a great job of editing papers and finding grammatical errors and things like that. It can write a lot of news stories. Certainly, reporting sports events or some of the more prototypical things that happen.

[00:27:57.27] And it can do a lot of patent search stuff and comparing patents and things like that. So you're talking about a lot of paralegal jobs, entry-level programming jobs, entry-level-- the cub reporter jobs, the bottom of the ladder for the editorial assistant jobs that-- this thing can do just fine. And that's a problem for those people who would have gotten those jobs and them not being out of work, either getting laid off or not being able to get the jobs in the first place.

[00:28:38.50] But it's also a problem, because at least right now, GPT-4 can do all the entry-level stuff. But it's not doing the really challenging, sophisticated, creative programming. The things that haven't been written before. It's not doing the work of a really superb editor who takes the novel and helps the author restructure it to bring certain characters more into focus and make the plot tighter and more exciting.

[00:29:09.51] And it's not going to do the investigative reporting that we still, occasionally, get, maybe, from The New York Times. I mean, so where are we going to find the people with those skills? How are going to train the next generation to do the more challenging level of those jobs if they're not starting at the entry-level? And what's that going to do to our society, if we don't have people who can program anymore and who can really improve somebody's writing. Are we just going to let GPT-4 do all of it?

[00:29:50.22] HARRY: There's a really interesting point. Any other follow ups on that? All right. So just shifting gears. A little bit like the first panel, we're going to speculate, but only two to three years out. So let me toss this over to Martha. I know you've been studying this for a while. I'd love to hear your thoughts about where this technology is going.

[00:30:13.08] MARTHA PALMER: Well, I've already mentioned the education application. My online chat application and the programming. It's certainly going to be doing lots of coding. And we

all know that being in Google or Microsoft and Google are about to go head-to-head on who's going to have the best search engine that's going to be the most interactive and do the best summarization and allow you to most quickly hone your search to exactly what you want it to.

[00:30:43.10] And I'm really looking forward to that. I think that's going to be a lot of fun for all of us. And that tool-- it may need, again, a little bit of tinkering, but it's going to-- could be just a fantastic boon to researchers everywhere. To historians, to classicists, to political scientists, to economists. I mean, the having a really sophisticated search engine that has, at its fingertips, everything that you want it to have read with respect to your field. Could be a game changer.

[00:31:22.31] And there's another-- but there's another-- there are a couple of other potential areas. And these are things that I think it probably could do. But I have a lot of reservations myself about whether or not we want it to do this. So when police spend a lot of time writing up reports about all of the interactions they have and everything that happens, theoretically, something like GPT-4, given the information about what they've been doing all day and a lot of it's on camera anyway and recorded, could write those reports for them.

[00:32:02.49] When you meet with your doctor, the doctors spend a lot of time writing up the report of your visit. A lot of that is semi-automated already. Again, GPT-4 could record the conversation and write that report. I mean, you'd want to have lots of guardrails around to make sure that the information is kept private.

[00:32:28.48] I've spent a lot of time doing research with electronic patient records and trying to get information about a particular disease and a particular treatment and trying to do that automatically. Certainly, a lot of that could be automated too. There's a lot of issues around patient privacy that have to be taken into consideration.

[00:32:54.13] And again, there's a human element in the police interactions. There's a human element in your interactions with your doctor. Are we really ready to just hand that over carte blanche to an AI and take out that human element? So I think that's scary. And then there's-- we had a wonderful panel right before lunch about talking about all the creative things that this technology can do. Certainly, it can produce songs, but with certain singers, personas, it can produce different kinds of images.

[00:33:33.70] It can write stories. And you could certainly have it, especially, if it was working in tandem with somebody, you could get a whole bunch of Star Trek episodes where it might not write the whole thing, but you could ask it for a few plots, you could pick the one you liked the best. You could then ask it to rewrite a few parts of that and get it to do more detail and do more dialogue. And next thing you know you'd have a whole episode of Star Trek.

[00:33:59.84] That might put a lot of people out of work too. People who are already having trouble getting jobs. But there's a whole other layer to this whole idea of having GPT-4 take over a huge part of our cultural artifacts. The production of our cultural artifacts, because artists, to a large extent, observe the world and changes in the world and then try to bring them to our attention and help them understand the way that society is evolving in either good ways or bad ways.

[00:34:39.26] And those cultural artifacts that our artists create can also impact our culture and our society. They can change the way we look at things. I was lucky enough to get to see Hamilton in Denver last year. And for me, that was a transformative event. I felt like it really changed the way I looked at, not just our colonial history, but what it can be like about interactions between people. It was really novel and creative and OK, you could say to GPT-4, "Write a play about a colonial figure and use rap as the musical genre" today.

[00:35:23.21] But I just don't think that would have the same transformative effect that Hamilton has had on so many people. Or something like the movie, Just Mercy. So if we end up relying on AI to create a lot of our culture and, of course, the AI is to a certain extent, just going to be regurgitating the stuff that's already seen, then how does our culture evolve? And what happens-- where do we get the insights into how the culture is evolving that help us understand what we're doing right and what we're doing wrong.

[00:35:59.89] HARRY: Oh. Those are really good points. And one theme that's come out again is, maybe, the worry that the humanity of a lot of processes, whether it's lawmaking or art or music making, will be lost. And we were never at this moment before, because previous versions of AI couldn't produce outputs that resembled the quality of human-level outputs.

[00:36:25.79] And now, I love your question, just because we can, should we? Like we should be at the moment we are thinking whether we should, not just whether we can. And I think it's a really important point. Eliana?

[00:36:37.38] ELIANA COLUNGA: Yeah. So I really like how you highlighted the artifacts produced by humans and how that impacts culture. And I think it connects to the two issues I pointed out before and how we can use one to solve the other, right? So hear me out. So if we have tools. There's a lot of things that are not accessible to me.

[00:37:09.20] I have visions of art that could be created, but not the skill to make them. I could make that happen with the right AI support, right? And so now, a human could use something with, maybe, very little training but with the right AI support. And then there's nothing sacred or preordained about a 40-hour work week. So with our extra time, maybe we have more time to consume these lovely, creative

products to produce these wonderful creations with the help of the AI. So I think-- I'm just envisioning a utopia.

[00:38:02.26] HARRY: Yeah. I love that point. I think in the early days of the internet, it was a top-down process where companies gave stuff to people, and then along came user-generated content around 2005. And it was because they unlocked all these tools that weren't previously available and people were able to go up on YouTube and be an instructor or what have you. So I love that idea of unlocking the creators in all of us, Orly.

[00:38:29.82] ORLY LOBEL: Yeah. Now, that's a great segue. So I'll put on my law professor hat when you ask about the next two to three years. I think it would be really critical to think about the role of policy. And following up on the how do we get to the better scenarios versus the worse ones, I think, it doesn't just happen. It happens through a policy process, a collaborative public, private redirection of some things.

[00:39:06.61] I want to plug that-- I think it was mentioned in the first panel that we would want a public option in some way of these technologies so that, what Harry describes as the user-generated stuff. The democratization of the creativity process, the invention process is accessible. And I think about that a lot.

[00:39:34.82] So I actually don't know enough, but I've been in this conversation now of, what is the end goal of some of these companies that are at the frontier of-- like Microsoft, Google, Open AI. What is their end goal when they're putting out a generative AI applications? And the way it's been explained to me-- I was just in a conversation with somebody from Qualcomm, and they were analogizing their technology that was, basically, chips that went into every other tech company that needed to use that and that was like how they standardize. And it was a B2B model.

[00:40:20.41] The argument is that this is where the money will be. And the goal is to, basically, replace with generative AI technologies, replace the App Store as it is today, and have these new platforms, basically, of application where you'll have the technology itself or the chat embedded and all these other uses that we've already actually heard in some way like case techs and others.

[00:40:52.36] And what that means is that we really need to pay attention to new forms of monopolies or anti-competitive structures and think about what are the frontiers of antitrust. These are big questions. But I think, again, in this context of law school and attorneys thinking about these things, I think that will be really important in the next few years.

[00:41:19.33] I'll just add that I think there's a lot of room. And again, this is what I want to see. My wish list in the next few years. And this is

what I feel like I've been arguing against. A tunnel vision of government where they're so focused on risks of AI that it's been stagnating the engagement and the integration, the deployment of existing technology for good.

[00:41:48.94] And specifically, in the work of government. So you mentioned like police reports and using generative AI to alleviate some of that. I actually think this is the game changer. That it's always been the case that like private industry is always ahead. And government regulation, policing, law enforcement, monitoring compliance.

[00:42:19.09] Administrative agencies, they're always playing catch up. We actually do have the data. I was just looking at-- some of my research, my earlier research has been about occupational safety and health. And I had written about how OSHA is so strapped for resources. Like, basically, has no real threat to companies. There's laws on the books, but companies don't really care about. They're not going to be inspected.

[00:42:47.66] And they're now like evidence, empirical evidence that if they were earlier-- if they were even adopting machine learning to just decide where to put their weight, where to have not just like random surprise inspections, but actually, and especially-- that would be the game changer in terms of, of course, less injury, less lives lost, but also just public resources and costs, savings. And the savings for businesses too of billions.

[00:43:24.38] I think that's really something that every function of-- so it's not just law enforcement and policing. And it's not just adjudication, which is also something that's really important. But in every turn, I've written in the equality machine about fighting trafficking and missing children. And an area that's near and dear to my heart is consumer protection and employee protection when there's all these unenforceable clauses and contracts.

[00:44:00.44] So rather than just thinking, oh, Chat GPT can draft a contract, also think about the FTC being able to use these technologies to detect unenforceable boilerplate clauses across the board, and for so many industries and companies. Again, that will be the game changer, I think. To match up the superpowers now that companies are using in their marketing and their for-profit activities to match that up with public capabilities and law enforcement.

[00:44:37.50] HARRY: Yeah. There was a really great point. I mean, this is an opportunity to make the government side operate much more effectively.

[00:44:45.26] ORLY LOBEL: Yeah. That's very utopian.

[00:44:46.65] HARRY: Yes. Yeah.

[00:44:48.08] ORLY LOBEL: You've said that out loud. I'm like, well, they do that.

[00:44:50.42] HARRY: Yeah. But I mean, there's the opportunity. And we'll talk about that shortly. But one thing we've heard is a lot of anxiety about the future. And as Viva brought up in the last panel, this always happens when we have a new technology that comes around. People worried about the photocopying machine destroying everything or the VCR. And we've seen in the past, the internet, car, electricity, social media. So there's always this worry.

[00:45:19.68] So I wonder, what can we learn from concerns of the past when a new technology has come along. And I heard this commentary from the journalist Ezra Klein, which I thought was pretty profound. And he said, "Think back to 1990, and then you say to somebody, 'look, what's going to happen. In 15 years, you can access all the world's information because of this thing called the internet. Information that used to take you days to find in the library, you can now find for seconds. You'll be able to communicate with anyone across the world for free, you will be able to collaborate. All the world's research will be available to you. And by the way, it'll all be in a supercomputer in your pocket.'"

[00:46:00.99] And you're like, "That's crazy." But half the world in 2005 would have said, there is going to be this utopia. By 2005, we won't be working, we'll be 20-hour workweeks. We'll have solved every disease. And then some other people would have said, "Oh, no. This global reach is going to allow these hostile people to come and pry into our lives and ruin it."

[00:46:24.99] And a little bit of both this happened, to be honest. But neither have we veered to the utopia nor the dystopia. And so let me throw this out to the panel. What can we learn from past interactions with major technological changes? Eliana.

[00:46:40.65] ELIANA COLUNGA: I think, systems have a way of self-perpetuating, right? So I think dramatic radical change is really hard. Because people don't want that in general. And specifically, people in power would not want that. And guess what? They have the power. So I think that's one reason you don't see, that utopia.

[00:47:15.63] So I think what we should do is threaten politician's jobs with AI, so we should all just give into the doomsday scenario where we have an AI overlord, and then let's see if there is some change in that work.

[00:47:32.44] HARRY: I don't think AI could make up facts enough.

[00:47:36.89] ORLY LOBEL: Because you said that we can ask you about this. And I would love to hear your response to this, because I've been really frustrated with that real doomsday scenario of like the

Strawberry Fields like, tell us a bot to produce as many Strawberry Fields and then they're going to kill all of us or other kinds of analogies. But what is your response to that? You said it's silly, and I agree, but--

[00:48:05.12] ELIANA COLUNGA: I think it's silly. I think it's a doomsday scenario that is very cultural-specific. And I think it is based on what we envision. Like, I am smart. A person who is smarter than me is just like me but more so. So you have people-- like, I literally saw a thinker-- what does he call himself? Anyway, I won't name names.

[00:48:32.45] But anyway, he really put this scenario about developing this thing that's going to kill everyone and put it in the water supply and thereby killing everyone. And then he added, that's what I would do. To which I-- like, what? No. No. Nobody should say that unless they're in some list. Yeah. That's not normal.

[00:49:00.30] And so why do we imagine that something that is super intelligent is going to want to dominate and destroy? That's a very colonizer point of view. That's not the only way of existing. You could have a super nurturing ideal, instead. You can have the most intelligent being in the world just wants everyone to be happy and get along and make cat videos.

[00:49:32.93] It doesn't have to be destroy and control. So I think, it just goes with a way of thinking about what intelligence is and about what power is for. And not everyone wants power for that reason. Not everyone wants power. And so yeah. that's--

[00:49:55.03] ORLY LOBEL: That's super interesting. In my research, I did compare why the Japanese and Koreans-- I'm heading to Tokyo actually on Monday. I'm a G7 representative to the digital transformation reporting to the World Economic Forum. And it was really rewarding to see how, in my research from my book I actually said that the Japanese have a very different historical relationship with technology, thinking about it as--

[00:50:30.29] Bringing it into their homes and kind of a savior. Helping the family, helping gender relations, and folding laundry and putting diapers on kids and teaching them and nurturing them and just being a companion. Whereas we have Hollywood that shows us know like the killer robot.

[00:50:57.47] And there's lots of reasons for this culturally, even spiritually, religion, and of course, you can't talk with that-- these are very flat two strong distinctions. But there is something to it where we see the policies that are coming out of these governments reflecting that. The EU is all about privacy and stopping things and, I think, the American reforms are very individual rights and not thinking about the distributive justice.

[00:51:33.45] So having this anti-discrimination. And this is something that's near and dear to my heart. In the employment context, we think about equality as anti-discrimination rather than substantive equality of actually redistributing resources and making sure that there's-- like here, it's called this bad name of affirmative action. It's before the Supreme Court if it's even lawful.

[00:51:59.36] But equality means that people get to be part of the conversation. So I absolutely subscribe to everything that you said that we have these cultural preconceived ideas about, if we're designing something that's super intelligent, will it have all these qualities that we associate with governing? And that's very specific.

[00:52:30.30] I would say on that, though, that one of the things that, again, I think is a responsibility of the public, of the public conversation, of the media, and also just the government, is to not only look at whether AI is trustworthy, whether technologies, like generative AIs. Where are the flaws and how do we redesign it?

[00:53:00.56] But actually also thinking about the human machine interaction and thinking about irrationalities in the way that humans understand machines themselves. So I mean, all of us, I think, talked about how there are capabilities these days that are clearly already outperforming human functions. And it turns out that even when they are, there there's a lot of mistrust.

[00:53:30.17] And that is also a cost. If a physician is getting a recommendation for treatment and they are told, this is experimental research that shows us, they're told that it was an AI that recommended it, they're more likely to override that. To insert themselves as the human in the loop, as we say. That's problematic, because it can actually be like the opposite of life saving. It can mean worse treatment. So we need to think about ourselves, how we think rationally, not just about having the best AI that we can--

[00:54:16.98] HARRY: Great. Those are all terrific points. And before we open it up to questions, let me ask just a quick question. And this was really inspired by Orly's book. We are not passive recipients of a predetermined technological future. So what active steps can all of us in this room do to shape the future that we want to see with AI? And I'll throw it out to Orly.

[00:54:42.72] ORLY LOBEL: Yeah. I'm smiling because, of course, I want everybody to read my book but I also want to say that Harry was one of the reviewers of the book in recent symposium that the Yale Journal of Regulation held. And you can find his article online. Do you remember the title, the exact title of

[00:55:04.53] HARRY: Something about Cognitive Bias--

[00:55:06.21] ORLY LOBEL: Yeah.

[00:55:06.72] HARRY: I don't remember my own article. Ask GPT.

[00:55:09.60] ORLY LOBEL: Ask GPT. Yeah. Yeah. But it was exactly on these points and really excellent. So yeah. I mean, let's get other comments.

[00:55:28.51] ELIANA COLUNGA: I think, one thing-- I don't know how much we can do this as individuals. But I think the solution, or we can go a long way by, basically, adding diversity at all levels. So not just who is creating these technologies, but who is dreaming them up, who is involved in the process of making decisions, who is testing them, who can have access to them, who are they used on.

[00:56:06.13] All of these things need to be distributed over more people. And so I think of issues of participatory research, participatory work. And I think that insight of how-- bringing to the table everyone who could possibly be impacted by it at every level of the process, I think, is really important.

[00:56:39.92] HARRY: That's a great point. Martha, could you--

[00:56:42.16] MARTHA PALMER: Yeah. So one of the things that became really clear under COVID when all the schools went online was how much easier it was for certain households to immediately get their kids online and on an iPad and doing homeschooling versus other households that didn't have internet access, didn't have the devices at home, didn't. And the Denver Public Schools made a point of making sure that every Denver household that had a child in school got an iPad so that-- I mean, they put a lot of effort into trying to make it as accessible as possible.

[00:57:19.23] And I think one of the things that I worry about with all of these advances is that it's just going to exacerbate the divide that we already have between the people who have easy access to all of this and the people who don't in this country, but just all around the world. So I don't know. I mean, if there's some way-- so Boulder's been trying to get this thing going where anybody who's doing any new building in town, even the commercial buildings, they have to put aside a certain amount of money for affordable housing.

[00:57:49.13] Because of that, we have so much trouble here with not having enough affordable housing. I'd love it if we could somehow get Microsoft and Google, when they're in the middle of their fighting, to somehow have to provide a certain percentage of the profits that they will surely to make this as widely accessible as possible. And maybe we can demand that. Maybe we can make that such a publicly important issue that they won't have any choice.

[00:58:22.94] ORLY LOBEL: And perhaps, this is too an obvious point, but everything that you said is, much, much worse in the developing world. And so I totally agree that we don't talk enough about how

there are millions of people who are not connected. We talk about all the dangers of our kids being online and all that. But we don't talk about the dangers of being left behind because we're not connected.

[00:58:53.80] HARRY: Yeah. Those are all phenomenal points. We need to be mindful of the distributional inequalities that this technology could even worsen. It would not be a good world in 20 years if only the rich get richer from this technology and those who have less we're left behind.

[00:59:14.43] MARTHA PALMER: And how are we going to stop the rich getting richer? They seem to be so good at that.

[00:59:18.14] HARRY: Well, They can Get richer, but not only the rich get richer. OK. Let me throw this out to a student for-- our first question if there's a student in the audience who, OK, fabulous.

[00:59:38.49] AUDIENCE: This is perfect, because this is actually where I want to take my question. It seems like at least in this country, for the last 50 years, things have actually gotten more unequal. And despite all of these-- well, different life expectancy, things have actually changed, and that's all complicated. But it's certainly from a purely economic perspective gotten more unequal over time.

[01:00:01.89] And trying to-- looking towards the more utopian side of things, like we're suggesting, is going to require a reshift of how we think about ourselves in society and our place in our relationship to work and all of those things. And I just can't help, but think about like, Kevin McCarthy and his debt ceiling negotiations is saying, we're going to get rid of SNAP. We're going to ruin these SNAP benefits. That's all going to be out the window.

[01:00:30.48] And it seems like part of what we're dancing around here is, in fact, a restructuring of how wealth is distributed and how power is distributed. And we can have this conversation in the abstract. But the kind of political implications of that seem to be like, we've seen this coming for a long time and yet we haven't actually acted on it before.

[01:00:56.49] Like the extended tax credits that were given out during COVID. It seemed like we were on the verge of a systemic shift, and then we went back. We're like oh, no, no, no, no. This expired. How do we take that momentum forward other than to be quite blunt waiting for my grandfather to die? But beyond that, what do we do?

[01:01:23.27] ORLY LOBEL: I'm probably the same age as your grandfather.

[01:01:25.49] [LAUGHTER]

[01:01:29.58] HARRY: I mean-- well, one clear answer is to get involved and vote yourself. And I'm sure you do, but a lot of people of the younger generation aren't as politically involved as they are. And you

can get involved in community, organizations, or local government. Things of that nature. I think that's really important. Orly, do-- you look like you're deep in thought there.

[01:01:59.36] ORLY LOBEL: I mean, these are really hard questions. It goes back to the trajectory of history of what you believe is human progress and the nature of humans and our political system. And do we have a better political system than-- like imperfect democracies that are like capitalist democracies. I mean, in a lot of ways, there's a lot of scary stuff going on in the world that is even worse than the rich getting richer, which is that there's a lot of shifts away from democracies.

[01:02:38.06] And that scares me a lot. So again, these are questions that go way beyond this question of where we are with generative AI and technology. But this is my optimist's you know streak where I do think that we can leverage, we can harness technologies that we're developing right now to-- a, shed light on the sources of inequality. That's really important where we talked about-- and you talked about how you can detect discrimination.

[01:03:17.39] You can detect with large data sets, like patterns of disparities and benefits that are being given. And infrastructure that's being invested differently in different communities. That's one thing. But we can even, more than that, see some of the root causes of some inequalities. Like, why is it that college applications are skewed or there's disparities there.

[01:03:53.03] Is it like biases in the way that the application is set up? There's lots of room for experimentation. And I think we should embrace that. How far does the inequities in investment go back, and how can we correct that? And I think that really, again, we need to keep this sense that there is a lot that AI can actually do in tackling the world's most wicked problems.

[01:04:25.75] So environmental issues. I mean, we heard in this morning that it's like this double-edged sword, of course. Like, they also take a lot of environmental resources. And so we need to have this public conversation about what is the level of usage that is-- it's a cost benefit issue. Like, we want to tackle-- some of the best solutions right now for climate change, for endangered species are really coming out of machine learning.

[01:05:01.65] To show how we can clean up the ocean. There's so many great programs like that are AI for good. And we need to look more meta on like, how do we look at all these AI for good functions and then also look at the cost of deploying them and have much more systemic analysis of what is the level of accuracy that we're comfortable with. Level of progress that we're comfortable with.

[01:05:38.61] In general, I do think, again, this is my optimist. I do think that even though we have the sense that equality or inequality has been deepening, that there's a lot of divide in wealth or deepening of wealth distribution. We also see-- when you see trajectories of human progress, we are seeing longer life expectancies, much more cures, much more accessibility to literacy and poverty alleviation.

[01:06:17.64] There is progress there to point to. And so this is just like, I think, what we need to-- beyond voting, like people who care like you and so many in the room about these things. We really need that skin in the game of, we have the technologies. Let's deploy them for good. That's, I think, the best that we can do.

[01:06:40.62] HARRY: Yeah. That's great. Those are really thoughtful points. Another question from the audience.

[01:07:00.54] AUDIENCE: This was a great panel. Thank you very much. I had a question about what proposals or gestures-- reforms to public governance or corporate governance you guys are most excited about and would recommend. And as a follow up, do you think our current administrative and legislative and regulatory tools are sufficient to the task of taming AI? And in particular, Orly, I know you mentioned the FTC. There's nothing like a platform technology, if not AI that seems to involve every type of commercial transaction we could possibly dream enough of.

[01:07:36.81] So I'm just curious if you all are, especially, bright eyed about proposals on the horizon for that type of infusion of democratic participation or at least control over this new technology.

[01:07:49.55] HARRY: I mean, just one quick reaction. I said on the first panel, I think we need something like an Apollo program where we, very quickly, have a lot of government funding to get the brightest minds in the room from all different sides and all the different perspectives and start figuring out the ethics, the governance, all these things, because, I think, nobody knows right now.

[01:08:11.82] And we need to figure this out quickly and do it well. And we don't want just a passive role in this and wind up down the wrong path. But I don't think anybody knows the answer to that, but I'll defer to the panel.

[01:08:28.98] MARTHA PALMER: So I'd say, right now, no. We don't have the right regulations and the right governance. But again, like Harry said, it's not obvious what that should be. It's going to be really tricky and really hard to figure out how to regulate it without stifling it and how to encourage the widest possible accessibility. So let's form a panel and figure it out.

[01:08:57.36] [LAUGHTER]

[01:08:59.02] ELIANA COLUNGA: I really like this idea of an Apollo project and massive investment. One thing that I think that we should think about and I worry about is too much of the lumping of the governance of AI as like a one thing, like one topic. Exactly they say, there's lots of different industries. There's so many different applications.

[01:09:23.55] And I'm not like a big fan of what the Biden administration did last year. Yeah. Just late last year, which is just have an AI bill of rights. To me, it's so abstract, because it's trying to catch everything that it's doing nothing. And the EU is similar. Like it has the EU AI draft, which just kind of talks in the abstract.

[01:09:53.37] Like if it's a high-risk AI, if it's a-- and I have not found that very fruitful. So in some ways, we need that Apollo massive concentrate. Like, let's think about this new technology as a thing that's unified. But in a lot of ways, we need every single agency to develop competencies to think about the applications in like the FDA and the FTC and the EOC and OSHA. All these different--

[01:10:25.06] Like I said before, agencies can actually really, very much, benefit from using these tools themselves and providing guidance to industry. And on that, I want to say that, this is what is very frustrating. There's very little of guidance to industry in a lot of things. So content moderation is a great example where we talk in very binary ways about you use the word governance, which I think is the right word to use.

[01:10:55.11] A lot of times, they use the word that-- we talk about regulation as like let's either have this continued complete immunity and user generate the content, or repeal Section 230, and have the liability shield or liability without everything in the middle of how do you govern things that are already done in the private industry. So every single platform does content moderation.

[01:11:27.22] So governance, really good governance is about finding the best practices, finding the standardization, collaborating with private industry and figuring out what works, what doesn't work, how do you differentiate between the snake oil, how do you create civil society that has these checks on the profit sector of-- I've invigorated my work similar to hackers that are professional third party civil society that finds vulnerabilities like cybersecurity vulnerabilities.

[01:12:04.41] Let's have these bias bounties of civil society that's finding all these bad applications of AI, and funding that. And having a lot more rules about what needs to be disclosed. Access to data, all these things. So that's what I think about as governance.

[01:12:25.88] HARRY: Great. Well, thank you so much. Please, join me in thanking this incredible panel for--

[01:12:31.36] [APPLAUSE]

[01:12:32.40] --phenomenal discussion.

[01:12:33.57] [APPLAUSE]

[01:12:37.64] So we've had a full day. We heard a lot of different perspectives. I think we learned a lot, but there's many more ideas to uncover and many more problems to solve. So while we're all thinking about this, please join me in our dessert reception in the next room where we can continue the conversation. So thank you so much for coming.

[01:12:59.93] [APPLAUSE]