Silicon Flatirons
KNOW WHAT'S NEXT.

Transcript

2024 Rothgerber Conference

AI and the Constitution

April 19, 2024

Silicon Flatirons and the Byron White Center are proud to partner together on a conference on AI and the Constitution. The conference merges the Silicon Flatirons annual Artificial Intelligence Conference with the White Center's annual Ira C. Rothgerber Jr. Conference on Constitutional Law. This joint conference examines emerging Constitutional issues implicated by the rapid advances in artificial intelligence.

- Timestamps correspond to videos published at: https://www.youtube.com/playlist?list=PLTAvIPZGMUXO1TspwnCAO4i73YKkfaUIf
- Visit the event archive at: https://siliconflatirons.org/events/rothgerber-conference-ai-and-the-constitution/ for more information on the event including the event brief, agenda, speaker profiles, session descriptions, and more.

Contents

Published June 24, 2024

Silicon Flatirons Center for Law, Policy, and Entrepreneurship
University of Colorado Law School
2450 Kittredge Loop DR, Boulder, CO 80309

Keynote: Professor Harry Surden

https://youtu.be/PWsb4BQ5r3M

[00:00:00.26] SUZETTE MALVEAUX: Great. All right, good morning. Welcome, everyone, to the AI and the Constitution conference. I'm Suzette Malveaux. I'm the Director of the Byron White Center for the Study of American Constitutional Law at the University of Colorado Law School. So thank you so much for being here in person and remotely. I'm excited to welcome not only those who are in the state of Colorado, but also those who are really all over the country. So thank you for being here.

[00:00:33.24] I first want to do a land acknowledgment and note that as we gather, we honor and acknowledge that the University of Colorado Law School is on the traditional territories and ancestral homelands of the Arapaho, Cheyenne, and Ute peoples. I also want to acknowledge that 48 contemporary tribal nations are historically tied to what we think of as the state of Colorado.

[00:01:03.28] So as part of the White Center, I'm really proud to be partnering with Silicon Flatirons. This conference merges the White Center's annual Rothgerber conference with the Silicon Flatirons' annual artificial intelligence conference. And so it's really a perfect blend of the two areas that both of these centers care very deeply about.

[00:01:31.48] So just a little bit about who we are. So the Byron White Center is premised on the belief that informed and engaged community is essential to our constitutional democracy. Our mission is to support excellence in constitutional legal scholarship, to offer opportunities to our students to promote justice, and to expand public knowledge and inform discussion of the Constitution.

[00:02:00.65] Silicon Flatirons also has its mission, and that's to initiate, sustain, and elevate the conversation about technology law, policy and entrepreneurship, and to inspire, prepare, and place students in these really important areas. Silicon Flatirons, I'm proud to say, was founded by our current attorney general, Phil Weiser, who is also the former Dean of the law school.

[00:02:30.47] So in terms of today's topic, when the White Center was brainstorming about what the topic of the conference should be this year, two things were really clear. First was the rapid advances of AI and how it was infiltrating our lives on every dimension. So you think about this tool and you think about this time. And so we were thinking about, this AI is everywhere, whether it's voting, surveillance, privacy, law enforcement, property rights, copyright, free speech, employment, you name it. So that's the first thing that came to mind.

[00:03:13.17] Second thing that came to mind was the erosion of fundamental constitutional rights that we have treasured or at least a threat to some of those rights. So it made sense to consider the role that AI can play and should play and maybe is already playing when it comes to protecting some of those most cherished rights.

[00:03:39.03] Now, we're going to explore that question in three different areas, and that's what our three panels are about. So we'll start off our first panel, AI and privacy. So that panel is going to explore the challenges in terms of protecting the right to privacy in the context of this AI explosion.

[00:03:58.89] So we'll be looking at, first of all, what do we even-- how do we even define, what do we even consider is sensitive information, so that we can protect it? We'll look at the need for data privacy protections that are tailored specifically to minority communities, right? If we think about guarding against exploitation, over-surveillance, and political deception.

[00:04:22.33] We'll look at some of the promises and also the dangers of law enforcement's use of facial recognition technology, and the efficacy of local laws that require impact assessments when dealing with algorithmic hiring. After that discussion of AI and privacy, we're going to turn to another question, and that really is exploring the opportunities and limitations when AI itself interprets the Constitution and legal documents.

[00:04:54.17] So for example, we'll look at some of the potential benefits, things like efficiency, consistency, perhaps less bias. We'll also look at some of the detriments or potential challenges, things like the elimination of discretion and the illusion of AI objectivity. So that panel is going to explore those questions related to legal interpretation, and really the role of human judgment, right, when it comes to interpretation of the law.

[00:05:25.17] And then our last panel is going to look at AI and AI speech and the First Amendment. So we know that the First Amendment has long protected human beings, right, in terms of listeners and speakers, and corporations as well to be honest, but what about AI?

[00:05:45.33] So we know that with the rapid advancement of AI since, basically, 2022 has definitely led to an era where AI-generated speech can match or even surpass human-generated content in terms of sophistication and substance. So that development has raised some pretty serious and complex questions about the First Amendment and AI speech, and really raises this important question about whether or not speech that's generated by machines is protected under the First Amendment. So we'll be taking a close look at some of the various implications of AI-generated speech and First Amendment jurisprudence.

[00:06:28.84] Now, a couple of things about the format for today. Just so you know, we are going from roughly 9:00 to about a little after 4. We are hoping that you can stay for all of the program. We'd love to have you. We get that people have competing demands on their time. So if you need to duck in and dip out, that is totally fine. We welcome you for whatever period of time. You can be here.

[00:06:53.66] We are going to have a keynote speech, and after that, we'll have our first panel. We'll take a lunch break around 11:45. We'll come back at 1, please come back at 1, and we'll have our other two panels. And then after that, we will all be invited to go and hang out at a reception shortly afterwards, right? So we'll have food and lunches free and the reception will be fun. So we welcome everybody to come and join us.

[00:07:22.42] For those of you who are participating remotely, I'm sure you know that your cameras or audio are off. But if you need some assistance in terms of tech support, go ahead and put your request in the chat and we'll try to get you some help. For all of you who have registered, which should be everybody here, you'll be getting an email shortly.

[00:07:42.77] And so that email will have a survey. We would really appreciate if you would take the time to fill it out, give us some feedback. A recording of the webinar, a recording of our programming, and then also some CLE instructions. So good news to all of my Colorado lawyers. You get five general CLE credits, and the students, you can receive care pledge credits. So you want to be on the lookout for that.

[00:08:10.28] So at this point in time, I would like to introduce our keynote speaker for today. I can't think of a better person for the keynote than my colleague Harry Surden. Harry is a professor of law here at the University of Colorado Law School. He is also the faculty director of Silicon Flatirons' artificial intelligence initiative. Not only do we get him, but Stanford is lucky enough to have him, where he is associate director of Stanford University's CodeX Center for Legal Informatics.

[00:08:47.20] Professor Surden is one of the leading scholars in AI and the law. He's a creator of computable contracts, and he has a number of articles that are seminal in this field. Many of you have probably already read some of his work, structural rights and privacy, machine learning and the law, AI and the law, an overview. These articles are of the crux of this field and have been largely cited, widely cited.

[00:09:17.74] Professor Surden has a background in computer science and the law. Before he entered academia, he worked as a professional software engineer at Cisco Systems and Bloomberg LP. He's a graduate of Stanford and Cornell, both with honors. And I am honored to be introducing him today as our keynote to kick us off. Thank you.

[00:09:41.44] [APPLAUSE]

[00:09:45.55] HARRY SURDEN: Well, thank you so much. It's such an honor to be here, and thank you to Professor Suzette Malveaux, who is a superstar in her own right in constitutional law. And thank you for that overly kind introduction. It's been just an honor to be organizing this with you, Suzette, as my co-host here. I also want to thank Dean Inniss, the dean of this law school who could not be here today, as well as Brad Bernthal, who is the executive director of the Silicon Flatirons who also could not be here today due to personal reasons.

[00:10:21.60] I also want to thank the amazing Silicon Flatirons team who helped organize this event, Nate Mariotti, Shannon Sturgeon Christine McCloskey, Sarah Schnittgrund, who worked incredibly hard to put this together. The incredible Lindley Bell from the law school, the events coordinator, Michelle Logan and Melanie Kaye from the Daniels Fund who helped organize yesterday's event, all the student volunteers.

[00:10:47.83] We have incredible experts from far and wide as far as Australia and as near as Central Boulder. And most of all, I want to thank you, the audience, for coming here today, both here and online. It's really an honor to be here today to give the keynote presentation.

[00:11:08.13] As Suzette, professor Malveaux said, I'm professor of law here at the University of Colorado and I also have this computer science and law background. I was a professional software engineer. So a lot of my research is on the intersection of artificial intelligence and law.

[00:11:25.62] So let me give you a little bit of an overview of my presentation today. We're going to start out by bringing us all up to speed on what artificial intelligence is and where we've been in the past. Then we're going to talk about what changed in 2022 that's gotten everybody all excited. We're going to look at the current capabilities of AI, the current state of the art as used in law today, and then finally, we're going to look at some of the implications for constitutional law and AI, the subject of this conference.

[00:11:58.18] So let's just start out from the beginning. I know many of you have a background in artificial intelligence, but many of you do not. So it's always helpful to bring us up to speed, make sure we're all on the same page. So I'm just going to start with an overview of artificial intelligence, what it is, and then we'll see its progression.

[00:12:16.78] So there's probably no one good definition of artificial intelligence that everyone would agree with. But one definition that I find useful is artificial intelligence, or AI, is using computers to solve problems, make predictions, answer questions, generate creative output, or make automated decisions or actions on tasks that, when humans do that, typically require intelligence.

[00:12:43.00] So what does that mean? Well, we're out there in the world. Humans are driving cars. We're playing chess. We're writing books. These are all activities that are associated with intelligence. So when we get a computer to do such a task that we normally associate with human intelligence, we refer to that as an artificial intelligence task.

[00:13:04.45] So that, of course, is not a particularly, a definition everyone will agree on because intelligence itself doesn't really have a widely agreed upon definition. But for our purposes, we'll think of intelligence as being associated with a series of higher order cognitive skills that are loosely grouped with intelligence. These include planning, problem-solving, reasoning, understanding and generation of language, learning, et cetera.

[00:13:35.62] So the important point to say is when we get a computer to do something that when humans do it are normally associated with intelligence, whether it's playing chess or identifying cat pictures or writing poems, computers do it very differently. They do not do it in the same way. So AI I think is largely a metaphor to human intelligence. We're getting functionally similar outputs from the computer, but through a very different process internally. And I want to emphasize that.

[00:14:07.71] So let me-- so this is something that many of you might have seen. This is ChatGPT using GPT 4. So let me just get a sense of the audience here. How many people have actually used ChatGPT before? So almost everybody. I'd say about 90%. How many of you use ChatGPT or something similar on a regular basis, every day or almost every day? So about 40%. And then how many of you regularly use a paid model like GPT 4, a frontier model? So maybe about 30%.

[00:14:44.09] OK, so that's helpful to know. So what we have here today is the interface for ChatGPT, which is the, it's a chat-based way of interacting with an underlying technology known as GPT, which we'll talk about in a second. You put something that's known as a prompt. A prompt is either an instruction or a question. It's a piece of text that the AI technology, the large language model responds to.

[00:15:11.45] So here I've asked ChatGPT to write a merger agreement. And the reason we would associate this as an artificial intelligence task, because when a lawyer writes a merger agreement, we generally associate it with intelligence or writing documents. So here it is. You can type it in there and ChatGPT can give you a pretty good first draft of a merger agreement, or just about any legal document, a patent. You name it, it can do it.

[00:15:44.29] So that's amazing and that is remarkable. So you might ask yourself, how did we get here? How did we get here where we have these general purpose systems that can do a pretty good albeit imperfect job at answering just about anything about any question or

topic you can ask? How did we get here with AI? So I think it's helpful to see our journey a little bit in the broader historical context of AI. So I'm going to bring us up to speed quickly.

[00:16:17.10] Roughly speaking, in the world of artificial intelligence, you can think about it not just in terms of tasks as we did before, which is the computer driving a car or playing chess or writing a merger agreement, but also according to, how do we actually build these systems? So in the history of AI, you can divide it into very two rough approaches. The first approach, which dominated from the beginning, we'll talk about is known as symbolic AI or knowledge representation.

[00:16:51.15] But the gist of it is, we, in the early days of AI, we built AI systems mostly by having computer scientists interact with domain experts, like lawyers or doctors, and manually inputting rules, right? So humans would create rules about the tax domain or medicine either in expert systems. And the overwhelming thing is that it's a top/down approach, where a human tries to see the world and implement it in computer code.

[00:17:25.83] There's a whole different approach known as machine learning, which has come to dominate the field since 2000, and machine learning is completely different. In machine learning, you give pattern recognition algorithms, lots of data, and the algorithms themselves learn the patterns and the rules from data. So it's much less about top/down manually entered rules and much more about the machine algorithm discovering the rules from large amounts of data.

[00:17:54.64] So if we want to see that timeline of artificial intelligence, AI goes all the way back to the 1950s. From about 1950 roughly to 1990s was the era of symbolic AI. That's where people were constructing rules, and much programming today, actually, is of this top down design. So if you use Microsoft Word or Excel, there are programmers in the Northwest, Pacific Northwest, typing a series of rules to make Microsoft Word print when you hit the Print button, for example.

[00:18:28.05] But this is the era of symbolic logic. And starting from about 1990 to about 2010 was the era of machine learning. Largely driven by the internet and the availability of large amounts of digital data, data in digital form, people started to realize we can use these pattern-matching algorithms to start doing interesting things.

[00:18:49.32] And in this period, we had self-driving cars. So when you design a self-driving car, you don't tell it, for example, if you see a stoplight, stop, right? So that would be a manually coded rule, rather you train these systems, use these machine learning systems, learn this after showing them lots of examples of good and bad driving over time and it begins to learn these rules on its own.

[00:19:18.58] So this is a really interesting era, 1990 to 2010. We'll call this the era of narrow machine learning. And that's because we say

narrow because these were good at very specific tasks that they were trained on. So a car that-- a self-driving car couldn't, for example, play chess because those are two different tasks. But AI machine was very good at playing chess and a self-driving car machine was very good at driving cars.

[00:19:48.63] What started to happen about 2011 was the era of deep learning. And this was largely driven by a few things, notably hardware improvements. A bunch of people started noticing that, hey, these graphics cards that I use to play these games with my computer are also pretty good at AI.

[00:20:06.76] And they revived this long dormant technology known as neural networks, which has been around for at least since the 1940s, and said, what if we brought this back and started implementing it in this new technology, in this era of hardware? And this is the era of deep learning. They rediscovered these techniques, these old techniques, using new hardware got better.

[00:20:31.72] And the era I want to focus on right now is the modern era where we are, because in this early deep learning era, all the way back to 2021, they started to do these large language models that could generate text. But as we will see, these were actually very bad. We're going to show examples of them. They weren't that good.

[00:20:52.91] So most people thought our general purpose AI systems are long far in the future. But there are also a bunch of algorithmic inventions, notably transformers and word vectors, which I'll talk about during the era, that set the stage for the modern day where we are today.

[00:21:12.39] So now we're in what I call the improved era of large language models. Things that existed maybe five years ago, which were not so good, have gotten good, very suddenly good in the sense that they're useful. So just putting this in a little bit of perspective, where we are, if this is the artificial intelligence techniques, we're on the left side here under machine learning. So we're in the era where we're learning patterns from data, rather than inputting rules manually.

[00:21:41.03] And then within here, we're in within the world of neural networks, which is this rediscovered technique which has started to work really well again. And then within that, we're in the era of deep learning, which I'll explain, but eventually, this involves taking this old technique of neural network and making them really big, scaling them up. So they used to be really tiny systems, these pattern matching systems, and now they're really huge. OK, and that's where we are today.

[00:22:10.62] So OK, so that's a intellectual history overview. Why are people so excited about AI? Wasn't there I 5 years ago and 10 years ago and 20 years ago? The answer is yes. But the difference is it's

gotten really, really good across a huge number of domains and it didn't used to be really, really good.

[00:22:29.30] So this is an AI and Constitution. I decided to undertake a very innocuous and unthreatening Amendment, the Third Amendment, which prohibits the government from putting soldiers in your house and forcing you to host them, right? So apparently that was a big problem around the Revolutionary War times.

[00:22:52.31] And then I asked it, I copied the text, and this is GPT 4, and I said, hey, can, under the Third Amendment, can the governor come and just force himself at my house, right? So I'm asking it to do a constitutional interpretation for a new thing that probably nobody's ever asked. And let's see what it says.

[00:23:15.03] And it does a pretty interesting analysis. It gives us the background of the Third Amendment. It tells us that this arose during Revolutionary War times when the British government was forcing soldiers on people's houses. But it says, no, while the governor probably cannot stay at your house without your permission, it's not because of the Third Amendment. The Third Amendment should be applied narrowly because it is specifically about soldiers, and the governor is not a soldier.

[00:23:47.81] OK, so it's doing an interpretation, which is interesting. So maybe we've solved con law here. We will see that in a moment. But the larger question of why this is so exciting is that AI systems from just two years ago could not do anything remotely like this, not remotely. And I'm going to show you examples, because we have time machines of AI systems from 2022, the distant past of two years ago, and we're going to see they were much, much, much worse.

[00:24:19.47] So this is a remarkable moment. This is why everyone's excited. You can just take just about anything, plug it into one of these advanced systems, and this is GPT 4, not the free version, which some of you use, but the more advanced version, and get pretty much a sensible, if not entirely accurate or authoritative answer. So what's interesting is that the answer is at least pertinent to what I've asked.

[00:24:48.30] So this is why people are excited. In just two years, AI made a leap from being OK and narrow to generally useful in huge domains and getting better. And we're going to see this-- I have some examples in a minute side by side. But before we get to that, I think it's important to understand the underlying technology behind GPT, right? So ChatGPT, as I said, is the chat-based interface. There's something under the hood called GPT, and I want you guys to just have, generally have an intuition for how this thing works.

[00:25:25.99] So the way we're going to do it is take the words GPT itself, look at what each of the letters stand for, and break that down and hopefully come out the other side with an understanding of how this technology works. So GPT stands for generative pre-trained

transformer. So we're going to break down each one of those words, understand what that means.

[00:25:51.89] So let's start with generative. Generative is a word used to describe a group of AI tasks that are associated with generating things normally associated with human creativity. So often when we think about human creativity, we think about humans creating art, music video, and text, books, movies, poems, things of that era.

[00:26:17.65] So that's important to emphasize because AI is used in lots of different areas that aren't like this. AI is used to predict things, predict the weather or drive cars. Those are not generative tasks. They're trying to do classification or prediction or other things. So when we're using AI to do something normally associated with producing human output, that's where we're talking about generative AI.

[00:26:42.52] So here's a picture of a generative art of Yoda in the style of the artist Edward Hopper. Here's another example. I've asked it to create Mona Lisa in the style of a Pixar character, and here it goes. It does it right before our eyes. So this is amazing. Normally, it would take a human to do this, but this is generative AI.

[00:27:07.27] It's also useful in music. So what you're about to hear is entirely AI-generated. This is from a new website called Udio. The vocals, the lyrics, the music, this is all fronted finished generated by AI.

[00:27:21.71] [AUDIO PLAYBACK]

[00:27:22.02] [MUSIC PLAYING]

[00:27:22.59] - (SINGING) Carolina, 5 foot 5

[00:27:25.26] Bright blue hair and big brown eyes

[00:27:28.47] See beneath the neon signs

[00:27:31.47] Of some old two-street town tonight

[00:27:34.62] We don't need to talk things through

[00:27:37.53] Well, you can cry the whole way too

[00:27:40.50] Whatever feels all right to you

[00:27:43.59] Cause you got baggage, I got room

[00:27:51.40] Singing Carolina, oh, it's a long and winding road

[00:27:57.28] [END PLAYBACK]

[00:27:57.70] HARRY SURDEN: OK, so to me, that's astonishing. Those vocals sounded like real humans. That melody was pretty good. That's a song I would listen to. So that's amazing. So this is generative AI. Completely started from a prompt, they generated that song. That's udio.com.

[00:28:17.35] OK, so why are we talking about all that stuff? Because GPT and ChatGPT is part of generative AI. It is generating something else, which is human text. So GPT is able to generate and understand human text. It uses AI to predict the next word, one word at a time, based upon what's come before.

[00:28:39.77] So that's amazing. AI, the way it currently works, literally, is just predicting one word at a time. That's why ChatGPT produces one word after the other, because it's just trying to predict, based upon your prompt, using probability to predict what the most likely next word.

[00:28:58.08] And while it appears to be answering your question, it's literally just predicting one word at a time based upon what you've asked the prompt, and what it, itself, has produced. If it has produced the capital of France is, it's going to make itself more likely to produce the next word Paris. So it's kind of remarkable. It's unreasonably good at what it does given its design. So that's a generative part. Let's talk about the pretrained.

[00:29:25.22] So this is the P in GPT. So pre-training is really important part of the process. This was a process where you take these neural networks and you unleash them on almost the entire internet, right? So you have these pattern detectors, these deep neural net pattern detectors. You release them on Wikipedia, my articles that are free for people to read, books, the New York Times, what have you, and it starts reading every single word out there and detecting the patterns of language. It detects the patterns of reasoning. It also is able to detect facts about the world.

[00:30:11.34] So it turns out when you release these pattern detectors, not on thousands or hundreds, but billions of documents and start to scale it up, it has these amazing properties of being able to learn, for example, that the capital of France, after reading Wikipedia pages and countless other articles, is Paris.

[00:30:33.20] So because these systems are trained on billions of words, books, things like that, and because they're so large in terms of the number of parameters, which are basically like the individual pattern detectors, they're known as large language models. So it used to be the case in a law school that LLM meant masters of law, but now it means large language models. But both involve training.

[00:30:58.40] So when you hear LLM, these are large language models, as opposed to the smaller ones which were, thank you for laughing at that joke, which are out there. OK, so that's the P. Pre-training is this incredible process that a lot of people thought, wouldn't work, where if you just release these systems on the internet and show them lots and lots of words, they somehow, magically, learn the English language and most other languages as well, as well as reasoning and facts about the world.

[00:31:34.38] And then finally, the T part stands for transformer. So this was basically an algorithmic innovation by Google in 2017 in a famous paper called "Attention is All You Need". I won't go into the technical details, although I have a paper about this, which I'll show you in a minute, but the gist of it is, this invention allowed AI systems to understand the context of what was being asked.

[00:32:02.52] So if you have a question what is the capital of France, the AI can look to the context words of what's being asked. So it's going to be able to see Paris and capital and be able to do associations and then come up with France. Previously, that, for various technical reasons, that was intractable, and Google solved this in 2017. And this technology, this architecture called the transformer has transformed or taken over most of AI these days and is in the basis of GPT and most other modern large language models and things like that.

[00:32:42.00] So the transformer-- so If you want to read more about it, I have this article that's hot off the presses just last week, where I explained how GPT and ChatGPT works in maybe excruciating detail. If you can't sleep at night and you want to fall asleep, read this.

[00:33:03.29] But just place in context, we have-- this is everything we've talked about before. So we've artificial intelligence, which is having computers doing tasks that are normally associated with human intelligence. We have machine learning, which is having those computers figure out the rules from data rather than manually putting in. We have neural networks, which is this pattern detection technology from the olden days which has made a comeback and is now really good thanks to modern hardware.

[00:33:31.58] We have deep learning, which is taking neural networks and making them really big. And then we have the transformer architecture. This is one way of doing deep learning, that Google came up with in 2017, that has transformed AI, and this is where we are today.

[00:33:50.60] OK, so if we look at-- it's important to look at the history of GPT to understand what happened in 2022. So as many of know, GPT 4, the current state of the art, was released by a company called, formerly a nonprofit, now a company called OpenAI run by Sam Altman. And they took Google's transformer innovation and made some crucial improvements on top of that.

[00:34:17.71] But it's important to note that ChatGPT was not always very good. So GPT 1 came out in 2018. It was bad. I played with it at the time. I was not impressed. GPT 2 came out. I was the tiniest bit impressed, but not that impressed. GPT 3 came out and I was a little bit impressed because you could write poems, but it did not look anything like AI. It couldn't answer questions. It wasn't being coherent. The first

time I was impressed, and I've been testing these for 15 years every time these things come out.

[00:34:49.44] ChatGPT in November 2022. This is the free version, GPT 3.5 that many of you used. And then we now have the most advanced version, GPT 4, which came out about a year ago in March 2023, although it's not the same model it was a year ago. It's gotten better since then.

[00:35:08.40] GPT 4, it's important to pay attention. This is the paid state of the art model. A lot of people make the mistake of, they'll see something and they'll go home and try the free version, and it's not that good, and then they'll walk away unimpressed. But they don't realize that the paid version, GPT 4, is actually significantly better. So it's important to understand that nuance.

[00:35:31.29] So let's talk about GPT 4. This is a huge advance in artificial intelligence in my opinion. And to appreciate it-- now, we all take for granted that AI can read the Constitution and do an interpretation, right? Maybe not. But I want to emphasize how, just two years ago, that was not even remotely possible.

[00:35:54.69] And the good-- what's interesting is that you can go back and find GPT 3 just from 2022, where it was released. It's online. It's still there. It's frozen like a time machine. So we can actually ask it the same questions to just appreciate how far we've come.

[00:36:14.03] So GPT 3, as I mentioned, didn't really look like AI, the way we think about it with ChatGPT. It's something that you could ask any question and maybe get back a sensible response. So let's look back, let's travel back in time, all the way two years ago, and see what the beginning of 2022, what AI was like.

[00:36:35.43] So for years, I've been testing these models because they've been criticized for not having common sense. So I would ask it questions that both a toddler would know, but also that are unlikely to have been written online. So one of the things about these systems, since they read the whole internet, they can simulate knowing things by just repeating.

[00:36:55.17] So if you asked it where to, these old systems, what do apples grow on, they'll know the answer trees because so many people have written that. But if you ask it something silly that a toddler would know and it's unlikely to have written, it would quickly expose its flaws. So here I asked this old model from January 22, how many legs does an Apple have and it confidently answered, an Apple has four legs, to which a toddler would laugh and say, what, apples don't have legs. So this was my common sense test.

[00:37:28.71] I also just went back and asked it the Fourth, sorry, the Third Amendment question that I just showed you and it answered things like, occupants, whether of private residences or otherwise,

have a right to protect themselves when disturbed by loud noises. What. So totally nonsensical. So this was the state of the art in just January 2022. So we were impressed by GPT 3 because it could produce lifelike text and produce poems and stories, but it wasn't able to answer and reason.

[00:38:05.99] So as you saw by the legal analysis, GPT 3 from 2022 did not appear to be a shining beacon of legal intellect. So this is why we were so shocked and this is why everyone like me who has been studying this field for a long time was amazed when GPT 3.5, the three version, came out in 2022, because it was the first system that could respond sensibly to almost anything you ask.

[00:38:33.13] And I mean sensibly, not 100% accurately. It was at least pertinent. So if you asked it about the Third Amendment, it would give you something relevant to the Third Amendment, if not accurate. And as you saw, the previous systems would not give you anything necessarily even pertinent to what you would ask it. So getting systems to be generally conversational about any topic was a long standing challenge in the field.

[00:38:58.05] The second thing that surprised everybody was that emergent properties of reasoning and problem-solving. So these things were trained to literally generate one word at a time based upon the words they had generated before by reading Wikipedia and a billion documents on the internet. They were not specifically trained to reason and problem solve, and this was an emergent property.

[00:39:23.16] So here's an example of GPT 4 solving a logic puzzle. It was not-- there's a logic puzzle I made up. It was not trained to do, this was not on the internet, and yet by reading lots of documents, some of which were about problem-solving, it learned reasoning and logic, and that was completely unexpected to people like me. So now we take it for granted. By the time, that was not, the fact that-- we predicted that it would be able to write text. We didn't predict that it could solve problems.

[00:39:58.88] And as I said, another interesting prediction was that it could answer arbitrary questions about anything. So the previous ones, you ask it a question and it might give you a sensible answer, it might not, right? Now, if you asked GPT 4, it can answer-- it can give you at least a coherent conversation just about anything. And that was not the case in the past.

[00:40:23.90] So I mentioned GPT 3 was released in November 2022 to great fanfare. So if you remember the timeline, the thing I showed you was GPT 3 from January 2022. Then just in November of that same year, we had GPT 3.5, which is ChatGPT, which could do all these things that GPT 3 couldn't do. Then just seven months later in March 2023, we had GPT 4, which was even smarter.

[00:40:52.24] So GPT 3, 4, and 5, the free version came out. It's pretty good, still had a lot of flaws. GPT 4 improved across every measure, including reasoning, problem-solving, language generation. Again, it's not perfect, but it's important to just appreciate the huge changes that happened in just a year.

[00:41:11.10] So this is the January 22 version of GPT 3. I asked it to write a motion to dismiss. That's its motion to dismiss. Now, just a few months later, I went and I found a random complaint on the internet. I copied the facts and I said, based on the facts of this complaint, write a motion, and then boom, off to the races. It can do it, and it can do this pretty, pretty well.

[00:41:37.52] And I'll talk about how well in a minute because I don't want to-- it does make mistakes, but what I'm trying to emphasize is the large leap. So the fact that it makes mistakes and is not perfect is besides the point. The emphasis is it went from not very good to pretty good in a really short time, and that's why people are exited.

[00:41:58.64] And the fact that this GPT happened in the area of language is what's interesting, because language as we know, is one of the foundations to modern communication, to knowledge, and to law. So if you can make something, as opposed to art and music are very important, but language foundational in a certain other way, and it allows us to build upon our knowledge base.

[00:42:22.79] So GPT 4, these large language models are foundational in a sense because they allow us to build new knowledge, do analysis of other things in the way other aspects of generative AI like generative music, which is cool, do not. So I will not hesitate to say that ChatGPT is one of the biggest breakthroughs in artificial intelligence in the last 20 years. I think it's amazing, but it is not perfect. So let us, of course, talk about the limits of GPT.

[00:42:55.39] This technology cannot fail-- it can fail. It is not perfect. It occasionally makes up facts, especially the older version. If you type in into the 3.5, it will hallucinate. It'll make up plausible sounding case names, as our friend the ChatGPT lawyer found out. GPT 4 is less likely to do this, but it's still not immune. Sometimes the information is out of date. It can be tricked. It can make reasoning errors.

[00:43:22.78] It can reveal private data, as Professor Ohm will remind us. There may be biases in the training data. We really don't understand how it produces the tokens, the words that it does. So there's kind of a lack of transparency and interpretability. So while amazing, there are really limits that matter which I'm going to talk about in a second.

[00:43:47.49] So where are we today? I just gave you the history. Today, we have some real competitors. The GPT 4. We have Claude 3 from Anthropic. We have Gemini Ultra from Google. We have some open source or open models. Llama 3, the open model just released

yesterday from Meta. AI keeps improving. I'll peer into the future in a minute. The next Llama 3 400B is currently in training, which is going to be the first open model that is competitive with the paid GPT 4. So that'll be interesting.

[00:44:23.95] OK, so we've seen it. AI can be used in law. How do we use it? So there's two main ways. One I don't recommend but I'm going to show you, and the other way, which is, if you're going to use it, this is probably the way to do it. The way I don't recommend is using GPT 4 directly by pasting in information.

[00:44:47.84] And the reason is, if you have a client privilege data or sensitive information, that's going to OpenAI, and unless you're part of their enterprise policy where they give you SOC-2 guarantees of security, privacy, and encryption, and that they won't train on it, otherwise, OpenAI says we will look at your stuff, or we could at least.

[00:45:13.30] By contrast, they have these other specialized legal systems, notably from Lexis Plus AI and also West/Thomson Reuters co-counsel. These are built on top of GPT 4. They use GPT 4 in the background. But they give you not only access to legal data, but security and privacy and confidentiality guarantees. So that's what I would recommend.

[00:45:36.35] Here's just an example of co-counsel from Casetext, from Thomson Reuters. So it's asking GPT 4 to do the analysis on the back end, but it's using its proprietary legal knowledge base to generate the answers. So it's much less likely to hallucinate. Similarly, Lexis Plus AI, I don't have a preference. I think they're both pretty good, provided you know how to use it.

[00:46:01.29] So I want to make a big caveat that there are huge limitations, and you could go vastly wrong if you don't use it correctly. But here is Lexis AI Plus answering questions and giving you Q&A about a recent Supreme Court case. And it does a really reliable job.

[00:46:22.43] GPT 4 itself, it, provided you don't use confidential information, is a quite reliable analyzer. It can analyze, for example, a torts fact pattern that I made up and does as well, if not better than any of my torts students. So it can give you a conclusion. This is just a slip and fall case for Walmart that I made up, and it gives an analysis and then a prediction about the outcome, which is in line with my own. Gave about a 30% chance of the plaintiff winning, which I won't get into the details, but that's a pretty good assessment.

[00:47:07.82] It can read contracts. One of my areas is computable contracts. So you can ask it questions about contracts. And pretty reliable provided you do it right, not perfect, though, but it shows the promise. Here it is answering scenarios in a homeowner's insurance policy and doing a pretty good job citing the actual provisions about whether something is covered or not.

[00:47:30.69] So let me tell you what it's fairly reliable in law today. The best use case is using it to summarize documents and asking questions, documents that you've pasted in, primarily, into the prompt for technical reasons as opposed to uploading it. I won't really go into the details, but I'll just show you what I mean.

[00:47:50.06] Here I'm literally-- here's a random Supreme Court case I found. I'm literally highlighting the whole thing, copying it, and then pasting it into GPT 4 and asking about it, as opposed to either asking about it generically and having it finding out, or uploading it and asking about the uploaded document. This is the most reliable way for strange technical reasons that I can go into in the break.

[00:48:15.26] But you can-- it's very reliable about summarizing documents. It's much less reliable about doing legal analysis, particularly in hard cases. This is why I want to emphasize. So I showed you examples of it doing interpretation, hard cases, but you want to be very, very careful. The way to think-- my mental model is GPT 4 is like an excellent, the top of the class third year law student, right?

[00:48:41.51] So third year law student can be really useful producing initial drafts, doing initial research. You probably wouldn't want to turn in the work product from a third year law student to the judge or a client. You want to double check it. You want to verify it. So that's how to think about it. But a third year law students are extremely capable and they can really give you a head start. They can provide a lot of productivity. So that's, I think, the mental model.

[00:49:09.41] All right, so let's talk about, in my last couple of minutes, given where we are, AI and constitutional law. So we've seen the frontier models like GPT 4 can read arbitrary legal texts and give you reasonable sounding analysis. So what do we do about that when we have documents like the Constitution?

[00:49:27.92] Is this a good idea to have GPT 4 or the future GPT 5 being used to interpret constitutional documents? Is this reliable? This is something we'll talk about in the second panel. So I invite you to stay. Do lawyers or judges or others properly understand its limits and its strengths? And I think the answer is no, but it does have some strengths which I want to talk about. But let's first highlight its limits.

[00:49:56.17] So remember we asked GPT 4, we pasted in the Third Amendment. We said can the governor of Colorado stay in my house under the Third Amendment and it said no. The governor can't stay there, but not because of the Third Amendment. The Third Amendment applies to soldiers. It says soldiers or the military, right? So OK, it gave me a confident interpretation.

[00:50:20.74] So we've solved AI interpretation. Well, not so fast. Let's ask Claude. Claude is another AI. So we ask Claude and it says, yeah, if you plug it into the Third Amendment, the Third Amendment means the governor cannot force you to house soldiers or other government

officials, according to Claude. Oh, wait, wait, so we have our first artificial intelligence circuit split here.

[00:50:50.52] So what I'm doing is trying to highlight the dangers of AI because it gives you the appearance that is confidently and reasonably answering your question and giving you the answer, and undoubtedly, judges out there are currently or either openly or secretly typing it in and they're getting an answer and they're like, Wow, AI is smart. I'm going to defer to it. That sounds reasonable.

[00:51:16.20] But in the background, it was doing a value interpretation. Does soldier apply just to soldiers, right, as it literally is or is it more broadly supposed to apply to forcible housing of government officials who are not? That's an interpretive question. Who knows, maybe I can do that, but right now, AI gives you the appearance of giving you solid objective answers when in fact that is not what's happening.

[00:51:41.91] So we need to talk about this today, we will talk about this today, and we want to have a conversation about to make sure that judges and lawyers understand these value choices that are secretly going on in the background non-transparently. We heard yesterday in our amazing AI and ethics talk about issues of training on past documents tends to replicate issues of the past, and that is a problem.

[00:52:08.33] We heard also about magnifying biases at scale. So if one judge might have a silly opinion, but if AI gives you a confident silly opinion and 1,000 judges are using that same AI system, that silly opinion gets magnified technologically. So there are lots of issues we need to talk about today.

[00:52:31.62] But let me leave you on an interesting positive note because I think there's a lot of good about AI. We, in academia, tend to focus on all the negatives, but it is not-- I'd say there are more positives than negatives with AI. And I'm going to peer into the future. Before I do that, let me give you a little invitation to be skeptical, even of me, about people that aim to predict the future, particularly on longer frames, particularly about AI.

[00:53:01.20] So a lot of people will predict out what's going to happen in AI in five years, and they'll also make predictions about what's going to happen with respect to society or jobs or things like that. And the reality is they don't know. So they don't know, I don't know. And we often are fooled by the confidence with which Sam Altman predicts a future of abundance or Elon Musk predicts a future of doom and gloom, but their confidence does not reflect the reality. They don't know and I don't know and you don't know. We're really bad at predicting the future.

[00:53:36.17] Well, the best we can do is make reasonable predictions in a narrow time frame, in the 1 to 2-year time frame, based upon current trends about the likely things that are about to happen. In my

view, that's the most reliable way to predict a very narrow slice of the future. And I invite you to, very much, be skeptical of those who predict broad impacts on society, and particularly beyond five years.

[00:54:04.92] But what is happening in this narrow slice based upon current trends? Large language model systems are getting larger. OpenAI is reportedly currently training GPT 5, the next generation, better than GPT 4. We don't know how better, but these systems tend to get better the bigger that they get.

[00:54:23.39] So GPT 3, the one in 2022 which was OK, was 175 billion parameters. That's big, but small compared to GPT 4, which was rumored to be about 10 times bigger, a trillion parameters. And GPT 5 might be 10 trillion parameters. Each one of those parameters means it's got more patterns to detect, more higher level concepts, et cetera. So these things will get better. We just don't know how much better nor what their impact on society will be.

[00:54:56.54] Other things that are happening. The original ChatGPT was trained on junky data sets, right? There's a lot of good stuff on the internet, but a lot of not so good stuff. In the early days, they trained it on everything. Now there's a lot of emphasis on training on higher quality data, which has been shown to dramatically improve the quality of the systems.

[00:55:18.08] There are new AI architectures and designs that are coming out. There's been a lot of innovation. I keep up to date with it and there's a lot moving. So remember, the transformer, that was 2017. There have been a lot of inventions on things like that. The hardware is getting faster NVIDIA just released a new graphics processing unit, which is the core of these, which has $200 billion transistors on a single chip.

[00:55:45.27] We're starting to see increases in AI agents or autonomous systems. This is where you say, hey, GPT, go out, do some research for me, and then come back, right? So it's out there or go buy me this thing on Amazon. So the system is making decisions on it's, a series of decisions on its own, rather than this one off back and forth thing.

[00:56:10.74] We're seeing longer context windows, meaning you can put more text in there. Right now, we're seeing up to a million, which is a lot of books you could put in there. We're seeing better planning, algorithmic improvements in law, in particular, we're seeing more reliability. We're seeing more verification. So we don't have the ChatGPT lawyer, right? We're seeing systems that, before they tell you, they'll check whether a case is real or not and not just hallucination.

[00:56:43.80] However, there's still limits, and you really need to know the limits in order to use these tools correctly. If you don't use them, understanding the limits, you will make mistakes. You will think that

you're getting answers when, in fact, you're getting subtle interpretations.

[00:56:59.23] So as our speakers from yesterday say, AI is bringing a lot of benefits. I think it's going to improve access to justice. It's going to improve the quality of the law. It's going to lead to better understanding, higher quality legal work for clients. But we should take this as a moment to improve the law as well.

[00:57:18.78] Use this moment to make AI and the law, to make the law more transparent than it's been, more equitable, less biased, allow more access to justice, make it fairer. This is a moment, I think, where we can make the law, the law that we want to see for all of us. So thank you very much. I appreciate your time. And if you want to learn more about GPT, you can read about it there. Thank you.

[00:57:45.90] [APPLAUSE]

AI and Privacy

https://youtu.be/wmafph7r-LE

[00:00:00.80] SUZETTE MALVEAUX: All right, so let's go ahead and get started. And in terms of the format, each of our panelists has been asked to give some thought to a particular question that pertains to their research. And then I'd like the other panelists to chime in. You'll have maybe 10 minutes to react to your colleague and be able to ask any questions of them or, in fact, weigh in on in the conversation.

[00:00:26.93] So let's get started with Professor Ohm. You've written a lot about the power of data inference and how it changes privacy law. And one of the examples I can give is an article you actually did about 15 years ago, which talks about how analysts can re-identify supposedly anonymized data. I'm wondering how the advances in AI over the last 15 years has maybe changed your thinking on the problem of inference, if at all? So go ahead and get us started.

[00:01:02.31] PAUL OHM: Great. And should we just stay in our seats? Is that simpler?

[00:01:05.13] SUZETTE MALVEAUX: It's totally up to you. You're welcome to come--

[00:01:07.52] PAUL OHM: I'm going to keep it here.

[00:01:08.70] SUZETTE MALVEAUX: OK.

[00:01:10.01] PAUL OHM: So thank you so much. I think the morning has already demonstrated that there is some magic being had today by bringing together these two kind of preeminent centers. For those who don't know, I spent my first nine years as a professor on this faculty. And I was one of the first faculty members that Attorney General Weiser hired to join the center.

[00:01:33.08] And now with the distance of time and I'm at a different place, I do recognize that Flatirons and the White Center are two of the preeminent centers in the country. And they do things on a different level. And they have more impact than almost any center I can name.

[00:01:48.42] And so the bringing together of them for the first time seems like an auspicious occasion. And it seems so appropriate, given how important these issues are. I will also note that I can't believe I'm the first person to want to call this the Flat White conference because it is a joke that writes itself.

[00:02:05.05] So that's what I'm going to think of this for the rest of the time. Newton Campbell is here from the land of the Flat White. And so it seems appropriate. So what I'm going to talk to you for 7 or 6.5 more minutes about is actually a general point. So I think it's appropriate to be the first panel discussion of the day that will frame hopefully some

of what is said on this panel, but maybe on other panels throughout the day, because I have been thinking throughout my career, if you want to crystallize or distill what I'm thinking, about arguments that recur time and again with the advance of technology as we think about, what is that technology doing to our legal rules?

[00:02:44.88] And so I will get to privacy as an example of this. But let me name one style of argument. And I would love to talk to the Colorado Law Review about possibly publishing this as a kind of small contribution. I think of this argument-- I need to come up with a pithy name. For now let's just call it the AI style of argument-- as an argument in three steps.

[00:03:07.00] Step one-- and this is going to sound familiar to all of you, no matter what field you're in-- the law, and sometimes really ancient law, focuses on a particular kind of fact pattern, a particular organizing principle, a central form of reasoning. That's step one. Step two is what Professor Surden's masterclass talk was all about.

[00:03:30.12] Technological advances put pressure on that fact, or that fact pattern, or that organizing principle, or that central reason. But it's step three that I really want us to focus on. Step three, many people say, well, given that the law has just been subjected to pressure by the technology, the law must yield. We need a new type of fact pattern.

[00:03:52.39] We need to reorganize the way we think about analyzing that type of legal question. So let me give you three quick examples. I'm going to end with the one I want to focus on for now. So think about intent. Intent is a well-worn kind of precept in many, many, many legal provisions. Many laws say that only human beings-- and that's really important-- who take certain actions, but not only actions, but with particular intent are civilly liable or criminally culpable.

[00:04:19.65] Just to put meat on those bones, the securities law says certain forms of intentional market manipulation are illegal. And intentional does a lot of work in cases like that. It's a defense to an investigation if you can say this was accidental. I didn't know as a human being that this is what I intended.

[00:04:37.12] However, now that artificial intelligence trades stocks, the question of whether they acted with the requisite intent, if they did all of the other things required for market manipulation, is either philosophically less than straightforward. And it might even be argued that it's completely nonsensical. So you have a smart stock trading bot. You say, go maximize profit.

[00:05:02.91] It may take steps that, if taken by a human, would clearly be illegal. But has the AI violated the law? Has the human who kind of trained the AI violated the law? So step three-- and step three, again, is what I'm hoping to push back on-- is let's rewrite market manipulation law. Let's find something other than intent to replace what the intent used to do in that role. And that just seems like the natural response.

[00:05:26.64] OK, let me give you a second one. And this one is familiar to what many people in this room have written about. We have a lot of laws that focus on not only the decisions that are made by a decision maker, but why they made those decisions. So many legal notions of fairness require the decision maker, when they hand out their decision, to say, here is why I am saying yes or no to your request.

[00:05:48.74] Professor Goodman, for example, has written about due process law. Due process law has this old set of elements attributed to judge friendly about what it means to make a fair decision. And one of the things in the very long list is a statement of reasons. That's step one.

[00:06:04.25] Step two, you can't do this with AI. AI is a black box system. It is an opaque decision maker. Many, many, many people in this room have written about the problems of interpretability and explainability, not just with neural networks, but with older forms of machine learning as well. So the hypothetical is state government has a computer program based on an algorithm.

[00:06:25.80] It is deciding who gets Medicaid benefits. Those who have denied benefits, they may never know why. They may never get an accurate explanation for that decision. Not even the designers of the system can give them an accurate explanation.

[00:06:38.34] Step three, let's redefine the constitutional definition of due process. It used to turn on this notion of giving reasons. We now have magic machines that cannot give reasons. So we rinse and repeat. OK, I'm an information privacy scholar, which means that I focus on constitutional and statutory rules that protect people from harm due to the collection, use, and sharing of certain forms of information about them.

[00:07:03.23] And many of these laws are premised entirely on defining categories of information, not all of them, but a lot of them. Let me give you three quick examples. HIPAA, the Health Information Portability and Accountability Act, it governs this critical definition health information. The Colorado Privacy Act, wonderful state privacy act. And full disclaimer, I helped write the regulations around this act.

[00:07:24.88] Although, I am not speaking here on behalf of anyone, probably not even myself. It defines a series of rules that one must take for all pieces of data, but particular special heightened rules for, quote unquote, "sensitive information", racial or ethnic origin-- I'm just going to read some of them off-- religious beliefs, mental and physical health condition, sex life, sexual orientation, and citizenship.

[00:07:46.58] And then if you want a constitutional example, because I do write quite a bit about the Fourth Amendment, the Fourth Amendment, of course, protects information for which we are entitled to a reasonable expectation of privacy. And the Carpenter case has

recently said that forms of data alone may support that kind of conclusion. That's step one.

[00:08:06.04] We define categories of information. Step two, one of AI's oldest known powers is that given fact A about a person, I can infer fact B. If I know what groceries you are buying, I probably know your religion. And I probably can make some very educated guesses about your health condition because of the foods that you're choosing and the foods that you are avoiding.

[00:08:31.50] And so because this is the step two pressure on law, I can infer that you have celiac disease. Should your shopping list be regulated by HIPAA because I can infer that you are an observant Muslim? Should the sensitive information rules of the Colorado Privacy Act apply to those rules? And then a well known example that's been debated a lot recently, because I can infer from your cell site location information that when you went to a particular place, you visited a Planned Parenthood location, does the Fourth Amendment protect that data?

[00:09:02.90] So now we know what step three looks like. Step three says, well, because of AI, everything now reveals everything. We need to throw out that old way of protecting privacy. We need to stop defining laws that treat some pieces of information different than other pieces of information. This is not a straw man article. I am writing a much longer article than the one-- Colorado Law Review would never want to go near this one. It's big, and sprawling, and messy.

[00:09:27.17] And in it I'm taking on one of the most respected, preeminent information privacy law scholars in the country, also a good friend of mine, your colleague, Daniel Solove. Solove has an article-- and even the title gets my blood pressure going-- called Data is What Data Does. And Dan argues, quite unpersuasively I might add, that due to this power of inference we just talked about, we should just stop writing privacy laws like this.

[00:09:51.78] We should get rid of the notion that we have health laws and that we have sensitive information laws. I am going to spend a lot of time explaining to my good friend, Dan, why he's wrong. But I just want to focus just on the nature of the step three move, which Dan is making.

[00:10:07.29] Embedded in step three is a value judgment about the malleability of law, the malleability of technology, and what happens when those two collide. I think that too often people making the step three move, and too often people in general, treat technology like it's the fixed part, the part we cannot control, and the part we cannot change. We almost naturalize it.

[00:10:30.94] We treat it like it's something that evolved. I'm so annoyed whenever someone talks about AI or technology evolving. It doesn't evolve. People create what it does. More defensively, some

people who are legal scholars might be saying, I don't know if technology can be changed. But I certainly can't change it. That's a little bit more defensible.

[00:10:52.36] That's about institutional competence. That's about letting people do what they do best. But more perniciously, there's just this technological determinism, the idea that even though, yeah, some humans are involved, technology just happens. It just marches. It just evolves.

[00:11:08.11] And on the flip side, I'm going to say less about, because you understand how we think of law. Law is something that we fight over. It's contested. We have politics. We have rules. We disagree. We chart the progress of it over time.

[00:11:20.19] And so it's no wonder that step three has so much allure. When you have the fixed deterministic vision of technology, like today's large language models, versus the pliable, and malleable, very social thing that we call law, well, of course, law has to yield to the imperatives of technology. And I just want to rebut that. I want to rebut that for sensitive information.

[00:11:42.58] But I want to say that, at the very least, let's double down on sensitive information laws. And let's build into our definition of sensitive information laws this power of inference. So the Colorado Privacy Act, again, I helped write the regulations. It's not just your religion. It's data that reveals your religion. I think that's a pretty neat fix, at least until the day when AI means that everything literally does reveal everything.

[00:12:07.51] But to end my talk, let's not stop with sensitive information. Rather than reshaping securities law that accommodate robots that manipulate markets, why don't we just ban robots from trading stock on the stock exchange? Rather than reshape due process rules so that we find something other than explanation to tether our due process rights to, maybe we just tell states they can't use AI to dole out Medicaid benefits if they don't understand why it's making decisions.

[00:12:31.75] I don't harbor illusions that all of these arguments are going to succeed. March of technology is super alluring. People treat it almost with this near-religious zeal. I just want to push back on arguments like these. Thank you.

[00:12:44.83] [APPLAUSE]

[00:12:50.55] SUZETTE MALVEAUX: Comments from your colleagues.

[00:12:55.71] SPENCER OVERTON: So as GW's a non-privacy person here, I certainly can't adequately defend Dan. I do have a question though in terms of your intent example, Paul. I am concerned, at least in the context of discrimination, whether it's constitutional interpretation, statutory interpretation, et cetera, about the heightened

focus on intent. If you look at, for example, interpretations of section two, which is really an effects test of the Voting Rights Act, where there's kind of almost a trend to cut away at Section 2.

[00:13:38.52] And so what does your take with regard to intent mean with regard to the future of AI and discrimination?

[00:13:48.88] PAUL OHM: Yeah, and I'll confess with apologies, I haven't read your recent contributions on this, which you've distributed recently. I'm also not an anti-discrimination expert. To be clear, I think there is a debate predating the rise of artificial intelligence about intent and anti-discrimination law versus just pure disparate impact.

[00:14:07.25] And so I don't mean for my intervention to have anything to say, much less disrupt that debate. Insofar as I've thought about it, I think having a rule that says something that has a very, very, very disparate impact should be a violation of the anti-discrimination laws, whether or not there's intent. So that's where my vote would be.

[00:14:28.47] My point is for laws where intent is doing some really useful rule that isn't so in debate, we shouldn't ditch intent just because the robots can't intend. So I don't think I would actually use that as an intervention in your debate. But it's really important for me in this essay, if I write it up, then I need to clarify that distinction. So thank you.

[00:14:46.69] CHRIS CHAMBERS GOODMAN: Thanks. Can you hear me? All right, thanks for your comments. And my question is addressing if you want to double down on the privacy laws and on sensitive information, then are we going to get in a situation where because certain groups aren't well represented in the data already and more data will be kept away from the machines, that they will actually become less accurate?

[00:15:13.63] PAUL OHM: Yeah, I mean, that's a much, much bigger and difficult fight we can have. And I feel like we'd agree on lots of it and disagree on lots of it. But it is true that privacy law sometimes when it operates well and sometimes when it is used cynically by people can be used to get in the way of things we want to do with data for sure, 100%. That's almost to me tautological, that sometimes if you really, really believe in privacy, it means understanding that there will be costs, including things that we really might care.

[00:15:43.12] So if a privacy law is standing in the way of making a biased machine learning algorithm less biased, I want to have that debate. That might be a great candidate for an exception. Or it might be that I'm able to persuade other people that the privacy interest should be stronger. But you're right.

[00:16:02.06] CHRIS CHAMBERS GOODMAN: But it's that balancing privacy and fairness.

[00:16:04.70] PAUL OHM: Yeah, I mean, being a privacy person means you're super unpopular among statisticians, librarians, historians, people who really believe that, like data is this kind of thing that we should never inhibit any access whatsoever to. And so I think it's the burden on all privacy scholars to make sure you're super nuanced about the harms you care about, the reshaping of society you're trying to prevent to justify what are really deeply felt beliefs on the other side about data.

[00:16:36.48] SCOTT SKINNER-THOMPSON: Thanks. Do we have--

[00:16:38.22] SUZETTE MALVEAUX: Any other comments?

[00:16:39.46] SCOTT SKINNER-THOMPSON: Yeah, I'd love to-- Paul, thank you. I mean, I totally agree with you that we need to make the tech yield, not make the law yield. But or and I'd like to hear a little bit more about how that plays out because I think that in some contexts-- and I'll talk about this a little bit later-- the starting point, the anchoring point for certain legal interventions to new technology has been moratoriums or bans.

[00:17:13.38] And that's generally where I align myself. But one thing I'm concerned about is that too often that may be the starting point. But we end up accepting what I would characterize as sort of Band-Aid regulations that in the process legitimize the technology and really don't address the myriad privacy harms and equitable harms that you're concerned about.

[00:17:37.89] And as one example of that, I mean, I think body cameras are an excellent example of that. Body cameras are, in my view, pernicious surveillance technology that have a disproportionate evidentiary surveillance impact on communities of color and queer people. And yet, they've been accepted. And we're like, oh, well, as long as access is regulated, as long as we control when the camera is on and off, the technology is acceptable.

[00:18:09.36] And I think that once we're in that debate, we've lost. It's legitimate. And now they're everywhere. And so I'm interested to hear more about how you envision us at step three calibrating those two.

[00:18:23.42] PAUL OHM: Yeah, and I'm going to focus specifically on sensitive, just because I've thought more about it. So I have an unsatisfying response, and then hopefully, a more satisfying response. The unsatisfying response is in my longer article on sensitive information, I really want to separate the question Dan takes on, which is, should we even treat some data differently than others, from the next obvious question, which is, OK, once we agree to do that, what extra rules do we put on the sensitive?

[00:18:49.65] And I feel like that's more your question. Is it a ban? Is it a moratorium? Is it a toothless risk assessment? Is it somewhere in between?

[00:18:56.83] And so I'm just defending the category. I want to have a longer fight then about that. But I'm totally sympathetic, I think, to what is the thrust of your question, which is it's not enough to build tiny guardrails around these technology. I'm completely dedicated in other parts of my agenda to thinking of democracy as a force that can redesign technology itself.

[00:19:19.96] I mean, it does anyway, whether or not we intend it. But I would love to see-- and let you first Amendment people on that last panel come at me about this-- I would love to see government rules and regulations that are intended to literally say you can't run a body cam technology that works like this, you have to add this, and force them to go back to the drawing board and redesign their products. So that's a little more of a direct response. Thank you. Thank you all. This was really helpful. Appreciate it.

[00:19:46.12] SUZETTE MALVEAUX: Thank you. All right, we're going to move on to the next speaker. I would like Professor Overton to take on the question, in the context of democracy, most popular discussions about AI involve deep fakes. Why is data privacy relevant to AI and elections?

[00:20:06.26] SPENCER OVERTON: OK, thank you so much, Professor Malveaux. And it is an honor to be here. Again, I'm not a privacy person. I'm a democracy and race person. I really didn't realize or appreciate privacy in this context until I wrote a couple of law review articles, one about platform liability for deceptive practices in the voting context, and this more recent one on the potential harms of AI for democracy in terms of racial harm.

[00:20:41.51] For example, in terms of last generation stuff, we think about social media, the business model, collecting data, building a profile on someone, delivering ads, and the Russians in 2016, basically creating pages and ads pretending to be African American talking about social justice. And then right before the election saying, hey, let's boycott this election. Let's stay home.

[00:21:15.16] And in terms of the Russians' ad spend even though Black folks are about 13% of the population, they accounted for about 38% of the ad spend by the Russians. And they were about 50% of the user impressions here in terms of the ads here. So now, new forms of AI, emerging forms of AI, and data can take this to a new level because generative AI is easy.

[00:21:48.33] It's easy to create more content, to generate more content quickly. You can create more persuasive ads that can be more customized to smaller audiences, audiences like Black males ages 18 to 25 in Philadelphia here. And in the future with current trends, we may get to the point where we're collecting individual data and designing ads, processing and designing ads tailored for the interests of particular individuals here, the bot of the Black Congressman who

has a video conference with the constituents and pretends to talk to them about their individual interests and know them, and then at the end says, hey, these Democrats are taking us for granted.

[00:22:40.64] We really do need to stay at home. That is a possibility. And it might not even be explicit. You could have an AI that kind of understands your interests and basically just gives you a lot of sports and entertainment and diverts you from thinking about politics and distracts you or a particular targeted audience here that's a possibility. Now, I know some of the big platforms have guidelines in terms of political advertising.

[00:23:11.57] I know the EU AI Act bans cognitive behavioral manipulation of people or specific vulnerable groups. But this feature is largely unregulated right now and this use of data in the US. Another example-- and Professor Skinner-Thompson's going to get into it more deeply-- is just this law enforcement, AI-powered tools that analyze cell phone data, social media feeds, closed circuit TV images, and facial recognition technology to basically surveil without a warrant Black Lives Matter protesters, dreamers, folks like that, and their allies and really kind of chilling speech.

[00:24:06.57] This is now low cost. So there are a lot of local law enforcement departments that can use it with very little oversight. And like I said, he's going to talk about it. I'm not going to get any deeper into that. Finally, another example, politicians using AI to create new voter restrictions. So we know that machine learning has been used.

[00:24:36.43] Algorithms have been used for years to draw gerrymander districts. But politicians have been pretty clumsy in terms of voting restrictions. So they basically say, hey, let's end early voting on Sunday. The souls to the polls and the Black church. It's pretty obvious what's going on here. It's kind of clumsy and obvious.

[00:25:00.71] But you could definitely see a Secretary of State whose party is supported by, let's say, 80% of whites and is supported by 60% of whites and opposed by 80% of people of color basically using machine learning and basically looking at voter files, not just voter performance, but also your supermarket bonus card, what you're buying when, your TV viewing habits, that kind of thing, and determining, hey, most white folks on Wednesday, they go out and they do their errands.

[00:25:41.98] And on Tuesdays, they stay at home and watch their favorite TV shows. And the reverse is true for people of color in terms of day. So we're going to move those late early voting hours that go from 5:00 to 8:00 in the evening. We're going to move this from the current day to a day where white folks are more likely to be out doing their errands and people of color are more likely to be home watching TV.

[00:26:10.82] And you do that. And in the next election, you see a slight uptick in terms of your voters and a slight decrease in terms of the voters who would vote against you. And your party has some benefits. And no one knows what's going on. It's not a souls to the polls type of Sunday voting type of thing. No one knows.

[00:26:35.09] So these are all very real possibilities here. Anita Allen has an essay on how data privacy uniquely impacts people of color generally. And I would just say that because race is the most significant factor in voting performance, more so than other demographic factors, and because politicians have an incentive to win an election, data privacy has unique implications for people of color in the context of democracy.

[00:27:09.67] SUZETTE MALVEAUX: Thank you. Comments?

[00:27:11.83] [APPLAUSE]

[00:27:17.02] CHRIS CHAMBERS GOODMAN: So I want to just push back just a little bit. You talked about the horrors for democracy. But what about AI for good in democracy? So one thing that we've been thinking about is, how do we get more people to choose to run for office, those who would be the politicians that would be more representative of our interests and those sorts of things? So how do we use AI to select before we get to the polls and have to elect from candidates that aren't of our choice?

[00:27:45.26] SPENCER OVERTON: And I'm going to do a piece for the Colorado Law Review on some of the benefits of AI generally. My last piece was on harms. I think specifically in this data privacy context, there are some good uses. One is mobilizing voters. Basically Black voters are maybe about 10 points behind white voters, maybe 5 to 10 behind white voters in terms of turnout with Latino and Asian-American voters.

[00:28:18.14] I'm talking about CVAP, Citizen Voting Age Folks. It's about 15% lower here. So basically, if you can use data to mobilize folks, I think that that's a good use to boost turnout, identifying the common interests of different communities. So if we have these kind of cultural markers that kind of divide us, but there are some real things that actually unite us, but they're not out front in identifying those things and bringing folks together, allowing politicians to hear from underrepresented perspectives, those are, I think, great uses.

[00:28:56.22] So definitely there are good uses of AI and data. And we've got to navigate this space to figure out how to maximize those good uses while also mitigating the harms.

[00:29:08.57] SUZETTE MALVEAUX: Great. And I just wanted to follow up with that because you mentioned the EU. And I'm wondering if you have some models that you're thinking of either coming out of the EU, coming out of history. You said this isn't the first time that this has

happened, or even states. Are there models that you would be attracted to?

[00:29:26.45] SPENCER OVERTON: I think that-- and you all may know some stuff that I don't know. But I think a lot of states have focused a lot on deepfakes. They focused on kind of deepfake porn, some of these issues here. In terms of using data-- you're going to know this better than I do-- to manipulate people, again, it seems like the EU, they basically, as you all know, they categorize uses. And they base regulation based on the use.

[00:29:59.34] So this psychological manipulation would be banned. Whereas let's say, a deepfake is just going to require disclosure and transparency. So this is their highest level of regulation in terms of a complete ban. And it doesn't seem like we're doing much about this. But you may know--

[00:30:22.49] PAUL OHM: It's a good segue into the question I wanted to ask you, which I think is another version of the benefits versus-- so a lot of the restrictions you're talking about in the GDPR and the AI Act are use restrictions. But they don't necessarily get in the way of the collection. But there are other parts of the GDPR that, for example, they're using against Facebook right now, which credibly might say to Facebook, you cannot do behavioral advertising anymore.

[00:30:47.10] You can't collect the information. And one thing that I've always seen-- trust me, this will become a question in the end-- is that I have colleagues who are really, really upset about micro-targeting for advertising. But they're also dedicated political junkies. And they love the efficiencies of knowing exactly where to target your last dollar and your donor's money.

[00:31:09.45] And so they don't want to throw out that baby with that bathwater. They want to somehow say, cut it out for ads, but keep doing it for campaigns. I'm much more extreme. I'm like, let's just dry up that market altogether. It'll make our elections less efficient. But it will have all of these downstream benefits.

[00:31:26.64] I suspect it will also take away from some of the benefits that Professor Goodman talked about. What do you think about that, a just flat ban on this sort of behavioral targeting? And you can't just say that will never happen.

[00:31:38.76] SPENCER OVERTON: No, I won't. I will say I really am nervous and concerned about good uses in terms of mobilizing this disparity in terms of voter mobilization and turnout. I mean, I think that there are some really good uses. I mean, if we think about things like who's susceptible to sickle cell.

[00:32:02.99] I mean, there are a variety of, I think, beneficial uses in terms of these tools. And simply because some folks are using them for

nefarious purposes, I don't know that we should should just outright ban them.

[00:32:22.38] SUZETTE MALVEAUX: Great. Other comments?

[00:32:24.58] SCOTT SKINNER-THOMPSON: Well, I would just add, I think this is super interesting. And I think I share your reaction. As the earlier colloquy with Professor Ohm illustrated, I'm deeply skeptical about most of these technologies. I'm also worried about law's ability to catch up.

[00:32:49.94] But in the context of certain technologies, including behavioral advertisements and the like, whether pernicious or advancing important democracy concerns, I actually have more faith in people to solve some of these problems than the law, which is to say that despite ChatGPT 4's intelligence, I still think we're pretty smart and savvy. And yes, the deepfakes and whatnot and the manipulation has an impact.

[00:33:26.15] There's no question. But I also think that we as consumers of technology and consumers of information, we have gotten pretty savvy too at realizing when we're being pushed, nudged, or cajoled in a particular way. And so this is one area where I think that with prolonged exposure, we as people using our intelligence will solve some of these problems.

[00:33:51.56] SPENCER OVERTON: So I think that that's probably right. I think my concern with that is just a take is it's these unsophisticated people, it's them, that they're not sufficiently digitally literate. And therefore, if they just put in more work, I think some of these things in terms of deepfakes, other things, I mean, when we talk about ad delivery and you don't know who's getting what, there's no way that someone understands what's behind it.

[00:34:27.66] So I think that you're right that we will become more sophisticated. But I don't think that that's a reason to say law does not have a role to contain or mitigate some of these challenges. I also think that, getting back to this other piece about outright restriction, this lower cost piece is a big deal. The fact that if I'm advertising, I don't have to spend a commercial in a metro area and try to hit everybody, but I can just kind of target and focus is huge.

[00:35:14.73] And I think that there are definitely some tough cases. For example, what if I bought an ad that basically said, hey, we're going to covertly reduce anti-Black bias. And we're not going to be blatant about it. We just know this ad that has nothing to do with race, it reduces anti-Black bias somehow. I think that that's a tough case in terms of people being manipulated, even though I would say it's a good purpose.

[00:35:42.02] So I definitely agree that there are tough, tough cases here. Finally, one concern I have-- and I'm shifting back-- is this notion

that we're all going to get to the point where we don't trust anything. And I know you're skeptical. I definitely get the body camera thing. But the person who's skeptical of the onlooker who records police violence because that might be a deep fake-- and normally, if I think it's real and I'm an average white person, I'm like, hey, I appreciate humanity and respect it.

[00:36:24.70] And let me go out to this Black Lives Matter rally. But if I think this is a fake, and I'm just being had because I'm sophisticated here to a point, then I might just say there are other things for me to do. That might be a deep fake. This might be somebody pulling my leg. I'm going to go home. So I am concerned about that development to a certain degree.

[00:36:47.71] SUZETTE MALVEAUX: Great. Thank you. Let's shift ourselves to Professor Skinner-Thompson. AI has already been deployed by the carceral system with significant implications for privacy rights, particularly for minority communities. Can you tell us a little bit more about those implications and some potential legal interventions?

[00:37:13.75] SCOTT SKINNER-THOMPSON: Yeah, absolutely. Just before I get started, thank you so much, Suzette and Harry, for your entire Byron White and Silicon Flatirons, Flat White, teams for bringing us all together today. I mean, on the theme of prediction, Paul sort of predicted the structure of my intervention.

[00:37:36.95] It's going to have those three steps, except for it's going to diverge at step three I guess, at least diverge from Dan Solove. So my view is that generally speaking, building on the theme of the conversation, as it pertains to most technological advances, not just artificial intelligence and machine learning, carceral systems have been able to exploit constitutional rules that were developed without those technologies in mind.

[00:38:06.84] And there are lots of examples of this over the last half century, where carceral systems have been able to use specifically lax interpretation of the Fourth Amendment. Such interpretive rules include the notion that there is no right to privacy in public or no right to privacy over information once that information is shared with another individual, sometimes referred to as a third party.

[00:38:35.52] And so examples where the court has applied these old rules to permit the use of new technology include, for example, law enforcement aerial surveillance of people's homes, even within the curtilage of their homes, and surveillance of their properties from telephone poles over time too. So it's not just one-off surveillance, but over long periods of time. There are, of course, counter examples.

[00:39:02.46] It's not like technology always wins. The court has said using infrared technology to look inside someone's home is impermissible. And there have been modest victories, as Professor

Ohm gestured to the Carpenter case, where the court required a warrant for law enforcement to access historical cell site information from service providers.

[00:39:30.54] But generally speaking, as Professor Andrew Ferguson and others have aptly described in detail, current Fourth Amendment constitutional doctrine is ill equipped to deal with law enforcement use of facial recognition technology. As Ferguson bluntly puts it, the Fourth Amendment will not save us from the privacy threat created by facial recognition surveillance. And this is particularly so under the current judicial political regime that is the Supreme Court.

[00:40:03.18] And my pessimism about this, I think, speaks, again, to the broader discussion that Paul and I were having at the outset is, given technological advances, the poverty of constitutional law, the poverty of democracy, and legislative and Democratic interventions, where are we going to look to? And I don't know that I have the answer to that.

[00:40:33.95] And so just to underscore, the lack of constitutional protections for privacy leave minoritized communities particularly vulnerable, given that many in those communities lack lived privacy, the precondition for legal privacy, and the fact that carceral systems-- surveillance is already predominantly targeting those communities. Moreover, as it pertains to facial recognition technology itself, as scholars such as Joy Buolamwini have noted, the facial recognition technology itself is built and trained on biased data sets that can lead to a disproportionate number of false positives for people of color.

[00:41:25.65] And in case you need more legitimization of that research, I mean, the NIST recently confirmed the same outcomes. So how is facial recognition being used by law enforcement? Well, disturbingly, as the Congressional Research Service explained a few years ago in its report on FRTs, quote, "The frequency and extent to which FRT is used at various phases of the criminal justice system is unknown," which could give us some pause.

[00:41:59.51] But some of the known uses are, the first is what's referred to as sort of identity verification. And if you have TSA Pre, and you live in Colorado, and you go to DIA, you've experienced this, which is where the TSA uses a face scan to ensure that the person presenting the identification is the person on the identification, so just verifying that the person holding the ID is that person.

[00:42:30.64] That is problematic in and of its own self. But there are much more troubling aspects, including matching photos of suspects-- so you don't have an identification-- to databases to identify the suspect. And those databases, some of them are built by law enforcement off of mugshot databases, which have its own bias, or driver's licenses, which is a huge data set.

[00:43:02.00] And some of them are commercially built by companies like Clearview, AI, just scraping social media posts. So sometimes it's used to match a suspect to large data sets to identify who the suspect is. But other times, it's just used-- and this relates to political surveillance and what Professor Overton was referencing in terms of Black Lives Matters protests and the like-- sometimes there's just dragnet surveillance of public space to see who's there and if they're a bad guy, and identifying people in real time.

[00:43:41.44] SPENCER OVERTON: Or if they're a trouble maker. It doesn't even have to be a bad guy. If you're against police or you're for defunding the police.

[00:43:48.08] SCOTT SKINNER-THOMPSON: Right. Bad guy, very inartfully [INAUDIBLE]. No, no, it's absolutely true. Well, and ti's predictive, too, because they're going to make you a bad guy. So we have this technology being deployed at scale by federal law enforcement agencies, by local and state law enforcement agencies. The constitution, unfortunately, is a weak tool to deal with this.

[00:44:25.42] And there have been legislative proposals to regulate it. But I think that those regulations, consistent with my view of how the body cam debate has played out, are simply insufficient. So some of these proposed limitations include limiting what are sometimes referred to as the probe photos, so the photo you're trying to identify who is in it, saying they can only be matched against mug shots. You can't match it against everybody that has a driver's license, that we need to make the training data more representative so we have less false positives.

[00:45:09.61] The mugshot databases should be scrubbed periodically so that people who committed crimes a long time ago aren't identified, that real time surveillance should be allowed. I mean, I think these are all good. But I think if implemented it will just serve a legitimation function and won't really limit the overall surveillance harm of this. And that's why I align myself with Professors Woody Hartzog, Evan Selinger, and those in Congress who have called for banning, or at least putting a moratorium on the use of this technology using law strongly to make the technology yield.

[00:45:55.49] So that's where I come out on this specific application at step three. And then just to link this back to Harry's chat a little bit, while I was talking, I asked-- I'm not paying for this stuff. So I just use ChapGPT 3.5, the free version. I asked them, is the use of facial recognition technology by law enforcement constitutional? And I've already been made redundant because it basically said what I said.

[00:46:26.43] It said the use of facial recognition technology by law enforcement raises significant constitutional and ethical questions. And opinions on its legality vary. Facial recognition technology raises concerns about privacy, due process, and potential bias. Critics say it

leads to unwarranted surveillance. And the inaccuracies are noted. Some cities and states have taken the steps to ban it.

[00:46:46.80] Ultimately, the legality of facial recognition technology in law enforcement contexts will likely be determined by legal challenges, legislation, and ongoing debates about privacy and civil liberties. I was paraphrasing there a little bit. But that's basically what it and I said.

[00:47:04.05] SUZETTE MALVEAUX: Good answer. I think you got it right. Spot on.

[00:47:07.57] [APPLAUSE]

[00:47:09.75] PAUL OHM: Sorry, who are we applauding for? I'm a little nervous.

[00:47:12.34] SCOTT SKINNER-THOMPSON: ChapGPT.

[00:47:15.33] SUZETTE MALVEAUX: Comments?

[00:47:19.43] CHRIS CHAMBERS GOODMAN: Oh yeah, I can start. So you mentioned that while it would be a good idea if we were able to make the data sets more diverse so that maybe it would be more accurate. But what about encouraging people of color to submit more to facial recognition software, which would in turn help in the long run by making it more accurate, but also then make us even more subject to surveillance?

[00:47:46.01] SCOTT SKINNER-THOMPSON: I mean, I think it's a great point. And it's another example of the privacy information trade offs. Sometimes being obscured is of real value. And sometimes it makes you more vulnerable. And I think they're ultimately value judgments.

[00:48:08.11] To the extent that the data, the systems, the FRT systems are coming up with false positives. That's deeply problematic. To the extent that they're coming up with false negatives for people of color, that's fabulous, meaning it's not working. So I do think that is a tension that has to be resolved.

[00:48:29.37] And Michele Gilman at Baltimore, I think, has written, in my view, among the most persuasively on these surveillance gaps for minoritized communities and how in certain contexts, particularly when it comes to government benefit programs and the like, that they need to be addressed.

[00:48:49.73] CHRIS CHAMBERS GOODMAN: Thanks.

[00:48:51.12] SPENCER OVERTON: So definitely none of us like the concept I think of errors that are racially disparate. And obviously, I'm very sympathetic in terms of the chilling here. But let me just put on my DOJ hat for a minute here and basically ask you, there is this reality of limited budgets in terms of local police forces and law enforcement and tax dollars.

[00:49:23.49] There is this issue of innovation. And just like other areas innovate, we expect innovation in terms of criminal justice. There's a question of many human witnesses are inaccurate in terms of their identification. And there's an argument that we could have technology that could be more accurate than some human witnesses here.

[00:49:51.12] And obviously, there are uses of technology, like DNA, that have exonerated some people who were wrongly convicted. So how do we allow for innovation in terms of technology in the criminal justice system? I definitely understand selectively focusing in on erroneous applications. But what are the guidelines so that we say we're not luddites? In the criminal justice context, there needs to be innovation and growth. But here are the guidelines here in this space.

[00:50:28.84] SCOTT SKINNER-THOMPSON: Yeah, thank you. I mean, I think it's a really difficult question. And I think I am sort of a Luddite at the end of the day. What I would say is I believe that before law enforcement is going to use the technology, there needs to be, at the very least, a persuasive case made that this is actually needed for law enforcement to fulfill its duties.

[00:50:55.33] And to me, most of the time in many of these examples, the technology is rolled out. Law enforcement is good at their job. I mean, crimes are-- and this is, again, I mean, assuming you're not an abolitionist, where I've aligned myself a lot, but assuming that there should be robust law enforcement, they've been catching people and imprisoning them with some alacrity for some time.

[00:51:29.95] We don't have a problem finding people and putting them in jail in this country. We do it more than anybody else. So I'm not convinced law enforcement needs a lot more tech to do that. And so if they're going to make the case that we need this tech to do our jobs, I think the burden has to then show that they're somehow missing something really problematic and they need this tech to do this.

[00:52:01.80] I will say in terms of use restrictions when it comes to things like facial recognition, one of the use restrictions that I am perhaps most sympathetic to is the idea that law enforcement could, for example, use real time video surveillance to stop an ongoing mass attack, but perhaps couldn't use that evidence in a trial for the purposes of punishment, but could use it as a means of thwarting the casualty.

[00:52:42.82] SPENCER OVERTON: [INAUDIBLE]

[00:52:43.61] SCOTT SKINNER-THOMPSON: Well, because I think saving lives is really important. But I'm worried about if the technology can use to-- and hopefully, there's other traditional evidence they can rely on to potentially pursue the carceral response if that's where we decide to go.

[00:53:04.25] SUZETTE MALVEAUX: Yeah, sure.

[00:53:05.90] PAUL OHM: This has been such a fun panel already. It makes me realize I need to get out more often because-- all of the privacy conferences I go to no one ever asks about useful data. We just assume that that's someone else's problem. So it's good to confront it.

[00:53:19.55] I'll also say for those who haven't read it yet, Brian Merchant's Blood in the Machine is a really lively retelling of the story of the Luddites and a comparison of it to the fight against big tech. And he frames the Luddites as people who were being abused by factory owners and who rose up and actually had some substantive wins.

[00:53:40.81] So yeah, I too am a proud Luddite. Scott, my quick question-- Neil McBurney and I were talking about this beforehand-- is in my understanding, NIST's latest report says they're still biased. But they're getting a lot better and that they're building new evaluation mechanism. And so I always worry about arguments that are based solely on the inaccuracy of machine learning because in the history of machine learning, almost all of the algorithms just get better with time.

[00:54:07.80] I almost feel like you need-- and you did, you gave us the fact that mug shots themselves are inherently biased and problematic. That to me is a strong argument that's going to last, as opposed to these algorithms are all biased. I think we may in the near future find out facial recognition is still problematic, but not biased anymore.

[00:54:26.38] SCOTT SKINNER-THOMPSON: I think it's a really critical point. And I totally agree.

[00:54:28.63] PAUL OHM: Yeah, OK. Cool.

[00:54:30.48] CHRIS CHAMBERS GOODMAN: I can just add one more thing. So you just brought up a point about if we let law enforcement use the technologies, but then prevent it from being used at trial, that ties in to Paul's step three. We change the law a little bit. We can change the laws of evidence, make that another exclusionary rule. And then we would still get the benefits of both.

[00:54:52.42] And that way, tech and the law would be interacting, both being manipulated a little bit. But would that be a good solution?

[00:55:01.69] SCOTT SKINNER-THOMPSON: Well, I think that's a use limitation that I'm open to. But again, I think one part of this is just negotiation anchoring. You start with moratorium. But I will say, too, I think that aside, we, because it's quite evident we don't know all the ways in which this is being used and the extent of its proliferation. But before we get to a concrete debate about use limitations, I think we need to just pause, and then have that debate, not have the debate decades after the technology is entrenched.

[00:55:51.28] CHRIS CHAMBERS GOODMAN: Thanks.

[00:55:55.01] SUZETTE MALVEAUX: OK, great. Our last question I want to go ahead and pose to Professor Goodman. How is AI being used in hiring? What are some of the privacy concerns that arise in that context?

[00:56:09.83] CHRIS CHAMBERS GOODMAN: Well, thank you very much, Suzette. And again, thank you all for having us here and participating in this event today. So some of the ways that artificial intelligence technologies are being used in the hiring process are in creating position ads. You can use ChatGPT to create the ad for you.

[00:56:26.85] And then that ad will be disseminated to particular people, again, using social media and all of those things. Once that ad is disseminated, well, then you can use artificial intelligence technologies to screen resumes, to call resumes, to figure out who are going to be the people that deserve an interview, for instance, the people who are able to get started. Once you do that, some companies are also using it to issue personality tests of some of the applicants.

[00:56:59.20] They want to see, what are their strengths and weaknesses? Is this someone who is likely to do harm to the company or someone who's likely to help out? If the artificial intelligence technologies are then used to conduct video interviews. And they're evaluating your facial expressions. I have lots of facial expressions and talk with my hands.

[00:57:18.85] I don't know what they think about that or how that would be evaluated. But that's another round of screening. So through all of these levels, we're using artificial intelligence technologies that we've just learned we don't understand. We don't know what they're doing. We don't know what's going on.

[00:57:34.27] So those are just some of the ways. So the concerns that arise from that is one that's already been mentioned. The data is retained. When I do that interview, guess what., two years from now if I apply to that company, they've still got that data. It's still out there. Well, what if I've gotten better?

[00:57:52.99] What if I've developed new skills? I'm already going to be tied down to the me I was two years ago. And I may not even get through the door. I may not even see the ad if the way that the artificial intelligence technologies created the ad was an ad that is not something women would like to see or is not something that my social media feed, which I hardly ever read, would provide to me.

[00:58:21.04] So I may not get the opportunity to apply for the job. If I do apply for the job, the fact that data is retained may disadvantage me in future hiring. Another concern is the three V's, the volume, the velocity, and the variety, how quickly everything manipulates, the volume of data that it will have from all of these interviews, all of these resume submissions, and cullings, and videos, and all that sort of thing.

[00:58:50.64] A fourth concern is the perpetuation of the status quo. We've got a closed loop AI is using the current employees to develop a job description, which then AI is targeting to particular people that it thinks would meet that job description, which then it evaluates to see if those people that it targeted with its job description created specifically for those type of people is actually right.

[00:59:17.47] So it becomes a closed loop. So if you don't want to diversify your company, if you want it to be a bunch of mini mes, just replicating the status quo, then it can be very effective at that. But unfortunately, that raises substantial concerns for people who don't fit in. A fourth point is the notion of cultural fit. A lot of companies are looking for someone who's going to fit in, who's going to blend in and not stick out as, whoa, who's that and where are they coming from?

[00:59:50.92] So the problem with cultural fit, though, is if you're using artificial intelligence technologies to help you figure out who's going to fit in your culture, those technologies are actually going to turn that into a rule. And cultural fit is going to become an actual qualification that it's looking for instead of the more amorphous content or amorphous way that we consider it.

[01:00:14.95] You go have a drink with them and see, do you get along? Can you talk about sports or something other than work? Those are some of the things that we think about in terms of cultural fit. So those are some of the concerns.

[01:00:30.24] Some of the potential solutions are to make sure there's more transparency about these processes and how they're being used. Have the vendors disclose to the employers this information. And also, just taking more time, I guess, to really figure out what's going on. So one example of an attempted solution-- and I heard Professor Ohm call it toothless. But I don't know if he was referring to this specific one-- is New York City.

[01:01:04.69] New York City recently adopted an ordinance that mandated four basic requirements for employers who are using artificial intelligence technologies in their hiring processes if they are hiring within the city of New York. And so the four things are you have to provide notice that the AI system is being used. You have to give job seekers an opt out chance.

[01:01:29.03] You also have to have an audit done, an assessment of the effectiveness. If there's any discrepancies based on race, or gender, or other characteristics, that needs to be a third party independent audit. And it has to be made public. Your audit results have to be made public.

[01:01:49.38] So that seemed like that was going to address all of these issues. It was going to really satisfy a lot of concerns. Some researchers at Cornell did a test. And they sent out students, college students, to go look at all these websites and look at all these ads, and

try to find the impact assessment, try to find the notice, try to find the opt-out option.

[01:02:18.25] And they had 30 minutes to try to do that for each website that they were looking at. And they found 5% of the information being disclosed. Most of them could not find the opt-out provision. Or they had to scroll through like 87 web pages, click on this, and then go here, and then go here, and go here. And then they could find the opt out position.

[01:02:42.01] So again, this was with about 1,000 employers they looked at. So there's more to consider. But I think that's part of the problem. So getting back to Professor Ohm's point about it being toothless, that's kind of what they concluded. While it's helpful because some employers are changing their practices, some of them are saying forget it. I'm not going to use AI technologies because I don't want to have to go through all these steps.

[01:03:05.79] Or we did our assessment. And it is biased, what we're using. So we don't want to report that out. But the problem is this notion of null compliance. And it was unclear whether or not employers were complying because there was so much discretion in the ordinance.

[01:03:27.88] The ordinance said if you are using these AI technologies and they play a substantial role in your hiring decision. So that's where the wiggle room was. Well, I don't think it's a substantial role. So I'm not going to post it. They also couldn't tell if maybe they had an impact assessment done. But it wasn't ready yet. Or maybe they hadn't found a third party independent auditor who was qualified to do it.

[01:03:52.33] So there were all these reasons why there was so low compliance. We don't know if all of the employers were actually not complying or if they were just on the road to complying, or felt that in their discretion they didn't need to include that information. So I think I'll stop there.

[01:04:09.53] SUZETTE MALVEAUX: OK, great.

[01:04:10.80] [APPLAUSE]

[01:04:16.09] SUZETTE MALVEAUX: Thank you. I wanted to start with a question of my own. And this piggybacks off of the conversation we had yesterday when we were talking about AI and bias. I know that there are a number of lawyers in the audience.

[01:04:27.61] So if you can speak a little bit to the lawyers and maybe people at law firms who are using AI to select their employees. What are some of the ethical rules? What are some of the professional rules of responsibility that people need to be looking out for so that they are not running afoul of those by using AI in an algorithmic hiring?

[01:04:49.17] CHRIS CHAMBERS GOODMAN: Yeah, so the Colorado Rule 8.4 of the Rules of Professional Conduct addresses this notion that you're not supposed to be in the operation of the practice of law discriminating. And so to the extent that you are aware now that I've told you and these, you know, today's events have explained about the potential for bias in using these technologies, that's something that attorneys really need to take a look at.

[01:05:19.25] The other things that I talked a little bit about yesterday in terms of the ethical rules or the duty of competence. If you're a lawyer and you're using artificial intelligence technologies, then you need to be competent in how you use them and making sure that you understand enough to know what you don't know and what those limitations are. We've already talked about the cost issue.

[01:05:42.30] You can minimize costs for clients by using these technologies. But you maximize costs if you get sanctioned by using ChatGPT 3 to answer your legal questions instead of 4. So those are a couple of the concerns. Was there more that you wanted me to discuss?

[01:05:58.00] SUZETTE MALVEAUX: No, thank you. I wanted just the benefit for our law students and those lawyers in the room to have a sense of what those rules are that we need to keep in mind. And we are going to take questions soon. So hang on. Let's have other people react to Professor Goodman's talk. And then we'll open up the floor.

[01:06:16.90] PAUL OHM: I think I'm just going to continue us on this what are the benefits versus these very real costs that I'm really worried about. I suspect a lot of this is because we've made applying for a job too easy. We've made it so frictionless through monster.com and Indeed that every job is now a national job. And you hear about minimum wage jobs that get 10,000 resumes.

[01:06:38.27] And so if that's your starting point, you have to use something automated. It would be deeply unfair if you didn't. But I wonder if that's the problem we should be fixing instead. Maybe there's something too old fashioned having a little bit of friction so that if you're going to apply for a job, it's because you really want that job.

[01:06:56.59] Now, I know there are probably impacts on people who are poor, and people who are busy, and people who have to take care of their children. And so we need to design that into the system. But I think a lot of the problems you're pointing to are because we have 10,000 applications.

[01:07:09.88] CHRIS CHAMBERS GOODMAN: Yeah, no, that's a really good point. And I think part of why we have that 10,000 applications is because workforces are trying to diversify and trying to reach out to other.

[01:07:18.00] PAUL OHM: Right. So it's all benefits and uses all the way down.

[01:07:20.79] CHRIS CHAMBERS GOODMAN: Yeah, and then you have a better shot of finding some more diverse candidates if you have a wider pool. But the other concern, though, is if we tried to go back to the more old fashioned way, where you research the company by going there and looking at their brochures and physically being there, ChatGPT can write you an amazing cover letter without you having to do any of that.

[01:07:42.86] PAUL OHM: Oh, God. All right. You've lost.

[01:07:43.91] CHRIS CHAMBERS GOODMAN: Yeah. So I--

[01:07:45.92] PAUL OHM: I will say I never do this. But the one word I haven't used today is friction. And a lot of my current work is thinking about how to build more friction into these frictionless systems, but in exactly the right place to bring about human values that are at a deficit.

[01:08:01.58] CHRIS CHAMBERS GOODMAN: Yeah, and so with some employers-- I mean, my son was recently applying for jobs. And he's a software engineer. And so they give you all these tests. They give you a test, and then a little bit later another test. And so if you really want the job, you keep taking these tests. If you don't, then you move on.

[01:08:16.92] But that is, again, a place where those with fewer resources or fewer options, perhaps more diverse candidates, would not make it through all of those rounds.

[01:08:30.01] SPENCER OVERTON: Chris and Paul, I think, Chris, New York seemed to be a lot about transparency when we talk about notice, and audits, et cetera. You talked about a lot of ways folks are using data in the employment context. Don't we just really need new norms when we talk about, how do we regulate inferences, how we use data. And what are those new norms?

[01:08:56.81] And then let me just press Paul on this. Don't we need new rules in this context, as opposed to just saying, hey, nothing's really new under the sun. We can just use the old rules. Don't we need new rules in this context?

[01:09:13.81] CHRIS CHAMBERS GOODMAN: Yeah, so we definitely need new norms. But so many employers are so self-interested and not interested in the public good. And so if you know that Chris Goodman applied for a job with your firm two years ago, you want to go back and look at that information if you still have it because that's going to help you make a better decision.

[01:09:36.96] So I think it would be fantastic if employers were more public minded and thinking about the privacy implications for employees and the fact that there is this unequal sort of bargaining power when you're applying for a job, as opposed to the person who

may give you the job. So I think that would be wonderful. But how do we get to that point?

[01:09:57.06] Well, maybe we get to that point because of this New York City example because some employers are saying, forget it, we're going to do it the old fashioned way because either this is biased, now we see it, or we just don't want to put it out there that we're using biased algorithms to do our job for us. So incremental change, maybe.

[01:10:18.88] PAUL OHM: Yeah, it's funny. I do feel like what you were describing was toothless. But it has such a non toothless response in the story you tell, where companies are walking away from it. So in a funny way, I've already outed--

[01:10:30.37] CHRIS CHAMBERS GOODMAN: It's dentures.

[01:10:31.13] PAUL OHM: I've already outed the privacy community as-- right. I've already added the privacy community as being a little bit of an echo chamber. But if you want to know the one thing almost every privacy scholar-- I'll let you speak for yourself, Scott-- agrees on, it's a kind of unified understanding and belief that any regiment built on giving people more information, and then making them choose is doomed to fail, that the companies who are regulated in those ways are so sophisticated at presenting the question to you at the worst possible moment.

[01:10:59.66] I just want to read this recipe because something's about to boil. Yes, I'll accept all your cookies. That's what makes proposals like this public enemy number one. And I have to do this. The standard citation for that is my good friend, Dan Solove, who's got an article on privacy self management, which is quite good. And I commend it to all of you. Maybe stop reading that for that.

[01:11:22.52] But your story has a twist for me in that a few companies walked away, which is more like a prescriptive rule, like we're going to make AI so burdensome in a funny way that you might just walk away and not use it. That's a rule. That's an interesting approach.

[01:11:36.97] CHRIS CHAMBERS GOODMAN: Well, it was a big adjustment when I moved to Switzerland, because the default there is no cookies, only strictly necessary is collected. And I kept saying, no, I'm going to choose. And then it was already selected for me that. So I was like, OK, I don't have to choose.

[01:11:51.73] PAUL OHM: You still the pop up, though. You still get the annoying pop up.

[01:11:53.17] CHRIS CHAMBERS GOODMAN: Yeah, you still get the pop up.

[01:11:54.28] SPENCER OVERTON: But isn't that a little lazy? We're going to just make it so hard for you, as opposed to grappling with

these issues. And what are the norms we're going to come up to regulate this?

[01:12:01.84] PAUL OHM: This isn't ideal. This isn't ideal. No, no. And that's part two. So the privacy committee that's given up on notice and choice, they've uniformly said, we just need to write rules. And we need to write rules, even though some will accuse us of being paternalistic because frankly, paternalism has embedded within it we kind of speak on behalf of the society about what we should be doing and not doing.

[01:12:22.70] So if that's paternalism, if that's the way you want to define it, then you kind of have to own that label a little bit. But yes, it's about actually deciding these things instead of coming up with procedures and nothing more.

[01:12:32.30] SPENCER OVERTON: I agree.

[01:12:34.09] SUZETTE MALVEAUX: Good OK, thank you so much. And I think we're going to--

[01:12:37.55] [APPLAUSE]

[01:12:41.31] We're going to open it up to questions. And it is our tradition to take a student question first. So in the back, I have a student way in the back. And if you would go and get us started. Thank you.

[01:12:55.77] AUDIENCE: Hi. Thanks for coming, everybody. My name is David. I'm a second year law student, quickly realizing that the most efficient way to a third year law student was paying for ChatGPT, apparently. My question is about how we're supposed to handle the fact that data is wielded in spaces that are designed to be manipulative.

[01:13:16.27] And I think the most prescient example of this is on social media platforms. Twitter makes me mad purposely through the algorithm so that I engage with it, for example. And I think the EU standard is one that came up of its banning things, like cognitive manipulation. But I get worried that that's pretty broad and pretty vague. Am I cognitively manipulated when I learn something on Twitter, for example, that I didn't know before?

[01:13:41.58] And I guess my question is, how do you see us handling that, whether that's step three turning to a legal reform or step three being a go to tech companies and change their reform? Who is the target that tries to address that problem?

[01:13:59.53] SCOTT SKINNER-THOMPSON: So I think the answer is all of the above. And I mean, this builds on the concept of friction, which is one that I've talked a little bit about and that I think Shoshana Zuboff uses in the end of her book as well. Absent robust command and

control rules, from a harm reduction standpoint, regulators can create friction for these companies that raise their operating costs.

[01:14:31.25] But you individuals can also throw them off your trail, not imperfectly. But you can confuse them. And this is Helen Nissenbaum and others have written about this in terms of obfuscation. And that involves a cost from the consumer. You have to do work.

[01:14:47.75] But there are things you can do to make how you are surveilled, tracked, and determined less accurate. And so I would encourage you to try to do those things. Don't be so predictable. And if you're really a true Luddite, like myself, get off Twitter. It's not enriching your life, in my view. I've spent plenty of years on it. And I'm happier since leaving.

[01:15:21.11] I know Paul is still on there.

[01:15:23.14] PAUL OHM: Barely, almost never. I've never been on Facebook.

[01:15:25.64] SCOTT SKINNER-THOMPSON: But that would be my view. Take the bull by the horns. Use your own engagement with these platforms to try to improve the quality of your life.

[01:15:38.59] PAUL OHM: So the one thing I'd say is there's an easier test case that's proving why your challenge is so hard. So for the last year, everyone's been obsessed with children. So let's be really literally paternalistic and try and protect our kids from body image problems and addiction.

[01:15:57.89] And so a bunch of states are experimenting with proposals to limit all of these things. And they are being fought and opposed by not only the people you would expect, but by a bunch of academics who are worried about First Amendment implications. And so if we can't even solve this for children-- and as a parent of a freshman in college, I see every day how the addiction has taken hold. And even though I tried to be a good parent, I failed in that regard.

[01:16:28.67] I would desperately love a law to protect children from some of these effects. And I don't even think we're going to get there because of the way the fight is unfolding.

[01:16:36.28] SPENCER OVERTON: In the democracy context, I don't know that there's a bright line. I think that we want politicians to engage with voters of color, for example, rather than suppress their votes in terms of winning. So I like folks trying to engage. So I don't think there's a bright line between political advocacy and kind of the stealth psychological manipulation.

[01:17:04.24] I do think that you can go too far. I think we've got to figure it out. But I think that what a lot of people don't appreciate, because we focus so much on this free market, everything's open thing, is that from the perspective of people of color who have

internalized experiences, like Indian boarding schools, or being prohibited from speaking Spanish at lunchtime in school, or Black women who have lost their job because they wear curly hair, this concept of manipulating folk with images really is a continuation of unfair cultural conquest that violates values of autonomy, and expression, and association that we say we purport to believe as a liberal democracy.

[01:17:57.21] And so I think we've just got to take this seriously.

[01:18:01.58] CHRIS CHAMBERS GOODMAN: If I could just jump in real quick.

[01:18:03.17] SUZETTE MALVEAUX: Go ahead.

[01:18:04.40] CHRIS CHAMBERS GOODMAN: Just one other thing to think about. We look at the EU as an example. But they don't have a First Amendment there. So that really is the difference.

[01:18:13.43] SUZETTE MALVEAUX: All right, we have a lot of questions, so I'm going to take three at a time, so you can get your question out, and then ask our panelists to go ahead and take them. So I have this gentleman here, and right here, and then in the way in the back. So those are the three, one right here, two right here, and third in the back. Please ask your question. And then I'll let the panelists respond.

[01:18:35.16] AUDIENCE: I'll ask my short question instead of the long question. So the quick question is, the AI auditing standards are really immature right now. But we're still starting to require AI auditing, as in New York. So what do we do about that in the meantime? How do we make sure the regulation actually keeps up with the status of what's happening as far as regulation, and benchmarking, and everything else to keep things safe?

[01:18:58.32] SUZETTE MALVEAUX: Thank you. The question in the back, can we have somebody-- thank you.

[01:19:06.23] AUDIENCE: Hi. So this conference is about AI and the Constitution. But I feel like much of the discussion on this panel has been about state and local level law enforcement, and regulations, and legislation. So realistically, what role do you see federal regulations playing in protecting these kinds of privacy, given how quickly corporations are developing new technologies? Do you think states should be leading the charge? Or are we just going to be subject to the Brussels effect?

[01:19:38.10] SUZETTE MALVEAUX: Thank you. And one more question right here.

[01:19:49.96] AUDIENCE: Hi. Thank you very much for the panel. I'll try to piggyback on the other questions. We had the Conference of World Affairs recently. And I think it's posted now. But one of the big

questions was about cyber security and all that sort of thing. And the audience generally was just flabbergasted about how insecure they are with their smartphones or their equipment.

[01:20:12.31] And it's almost like going back to the old school stuff, a flip phone, to really get around a lot of those problems. But my question is basically is tax off of ethics. And speaking with Harry, it's just a question of quantum computing and nation states. And it's the supercomputers and so on and so forth. So it's just battle going on with information. The information wars are going on.

[01:20:40.72] And at some point, speaking with Harry after his lecture was-- I don't know how to phrase it. But at some point, with this gentleman was saying was about monitoring and how do you keep on top with ethics involved? And Paul Ohm was kind of getting into that as far as the whole ethics and HIPAA. So it's complicated. But it's simple. At some point, is it the Hollywood Terminator, the machine's ultimately going to be for efficiency or whatever the right word is, editing out the ethics. And it was great the example that was provided with doing the statutes and doing the case law.

[01:21:22.12] It's pretty good for AI to do that. But where do you bring in the morality? And again, judges are sworn in on a Bible, not the Rules of Civil Procedure or something like that. So thanks very much. Appreciate it.

[01:21:35.33] SUZETTE MALVEAUX: We've got three questions on the table. Who wants to take what?

[01:21:37.91] CHRIS CHAMBERS GOODMAN: OK, I can do a little bit on the first two. So in terms of auditing and assessments, keeping up with the changes in technology, and what is going on on the federal front, there was recently, a couple of weeks ago, an announcement about the Office of Management and Budget new policy that is going to require when the federal government is using artificial intelligence technology.

[01:22:04.62] So any parts of the federal government, any agencies, they have to make an assessment of, does this impact people's rights and/or does it impact people's safety? And if it does, then they're required, again, to implement these concrete safeguards. They have to have the risk assessment done. And the directive says, if you cannot adopt appropriate safeguards in this circumstance, you must cease using the AI systems.

[01:22:31.83] It requires a catalog of all the systems that all the agencies, all the federal agencies, are using. And that also will list, what are the vulnerabilities? What are the concerns? What are the risks? What are the harms? So that is on those two fronts. So I think that will help because it's not telling you exactly what you have to do. It's based on the risks of harm and what kind of rights are being potentially impacted.

[01:22:58.99] SUZETTE MALVEAUX: Thank you. Other responses?

[01:23:01.99] PAUL OHM: On the second question, I think it should be all of the above. I actually just hope Europe saves us from ourselves partly because they don't have to contend with our weaponized First Amendment. I mean, it isn't AI and the Constitution Conference. The phrase digital Lochner is often used to describe the way the First Amendment has been argued for, not yet embraced.

[01:23:23.87] But the Supreme Court will have plenty of opportunities to embrace that. I actually think-- and I'm biased because I worked on it-- but the state of Colorado's law is extremely significant in the fact that the legislature continues to be competitive with California. There's a healthy competition now among the states.

[01:23:42.91] And so I just read the headline yesterday in The New York Times that Colorado has added brain scans to the CPA, which I think is great. And then last but not least, you said something about the Federal government writing. I don't know what that is. There's another thing that writes statutes? I've never heard of this one.

[01:24:00.96] And then lastly-- I could go on for hours about ethics. This is something I'm thinking a lot about with the rise of generative AI. Let me just give you the teaser. I think we're going to learn that we don't need the technologists to be the only ethical people. That's been our approach for the last 20 years.

[01:24:15.63] We're going to be able to stick other people other than technologists into these companies. And they can do the heavy lifting on ethics, which I think would be great, if we could pull that off.

[01:24:27.67] SPENCER OVERTON: So in terms of the state or federal regulation piece, I think that state, local, and federal entities have frankly just-- there are constitutional problems with all three in terms of the issues that Professor Scott Skinner-Thompson was talking about here.

[01:24:51.88] And so I think that this court may not recognize them. But I do think that there are constitutional issues in terms of all three. Right now, very frankly, our federal government is broken in terms of lawmaking. States definitely have a role in terms of getting new ideas out there.

[01:25:12.73] We had a conversation yesterday after the panel about this. And hopefully, if we have some different states doing some different things, industry will come to the table with academics and public interest folks and say, hey, we need kind of one standard. We need some consistency. We're actually willing to come to the table and compromise so that we can have some uniformity and some clarity.

[01:25:41.08] And we can get something done. Finally, I'd just like to say we've had these instances before in terms of innovation and money, cotton gin. Cotton was over half of our exports until 1860 and

our leading export until 1932. But obviously, there were a lot of costs with cotton in terms that we're still dealing with in terms of racial equity.

[01:26:09.32] Similarly, Industrial Revolution is great. It's the reason we're in this room with these lights in this comfortable environment. But we're dealing with climate issues because the costs were externalized to other folks. And law did not step up to the plate here.

[01:26:26.42] And so for folks who are law students, I think they're well positioned to build careers to ensure that as this technology emerges, certainly we can have some innovation and some good things come of it. But those harms won't be unfairly externalized onto others so that we're kind of dealing with these large societal problems that we don't even know if we're going to be able to contain.

[01:26:57.17] SUZETTE MALVEAUX: Thank you so much. I think we all need to be at the table, including us Luddite. So I appreciate the shout out to Luddites being one of them. I am sorry that we don't have time for all of your questions. The good news is we have a lunch break.

[01:27:11.07] And so for those of you who I wasn't able to call on, please feel free to address your questions to our panelists during our lunch break. I'm asking you to go to the second floor. And lunch is held in Schaden Commons. So go ahead and proceed to Schaden Commons. There'll be plenty of time to talk more about this topic. And we're going to resume at 1:00 here. So please come back at 1:00. Thank you so much.

[01:27:38.99] [APPLAUSE]

AI and the Interpretation of the U.S. Constitution and Other Legal Documents

https://youtu.be/zMulrlANFZ4

[00:00:01.51] FYNN FEHRENBACH: Welcome back, everyone. I hope you all enjoyed your lunch. Yeah, so my name is Finn Fehrenbach. I'm a first-year student here at Colorado Law School. And I'm excited to introduce the next panel, AI in the interpretation of the US constitution and other legal documents. Here panelists will explore the potential risks and promising benefits of using artificial intelligence in the context of legal document interpretation.

[00:00:25.39] I'm super happy to announce that the panel will be moderated by, Professor Harry Surden, an expert in artificial intelligence in the law and the faculty director of the Silicon Flatirons Artificial Intelligence Initiative. Without further ado, I'll pass it over to Harry Surden.

[00:00:39.19] HARRY SURDEN: Thank you so much Finn for that kind introduction.

[00:00:41.55] [APPLAUSE]

[00:00:43.76] So I think we've got a really fascinating and timely, and maybe one of the only panels of its type in the country on using AI to interpret legal documents such as the US Constitution or other documents such as contracts. What I'm going to do is introduce our all-star panel, then give a little bit of overview before opening up questions to the panel. But let me start going in just in order here.

[00:01:15.18] And these are very accomplished panelists. But in the interest of time, I'll only provide very short biographies. But I encourage you to look at the more extensive accomplishments which are linked online and also in the program. To my left is Yonathan Arbel, who is silver associate professor of law at the University of Alabama Law. Has done a lot of really interesting work on AI and contracts, and is a well renowned scholar.

[00:01:42.38] Next to Yonathan is Megan Ma. Megan Ma comes from Stanford University at the Stanford CodeX Center for Legal Informatics. Megan is an expert in AI and linguistics and one of the leading thinkers in this field. Next to Megan is Vivek Krishnamurthy. Vivek thankfully has joined our faculty this past year. Yes we're very grateful. Vivek has done really pioneering work in the world of technology and human rights and directs our Technology Policy Clinic.

[00:02:17.30] And last but not least is Andrew Coan. Andrew Coan is not only the assistant associate dean for research at the University of Arizona, he's also an eminent professor of constitutional law where he writes a lot about constitutional theory. So we're really honored to have this distinguished group of panelists here today. So what is this panel about?

[00:02:43.34] Well, if you were unfortunate enough or fortunate enough to see my keynote, you saw that we're now in a moment where large language models like GPT-4 can interpret any legal document you give it and give you an answer. So we saw in that panel we took the Third Amendment of the US Constitution, and we asked it various questions about whether which the Third Amendment on its phase bars the government from quartering soldiers in your house? Forcing you to host them.

[00:03:16.28] And then I asked them, well, I asked GPT-4 would the Third Amendment bar the Governor of Colorado from being forced to be quartered at my house? And GPT-4 said, no. The Third Amendment gave me an interpretation. Only applies to soldiers. But then I asked another AI system another frontier model, Claude-3-Opus, and I asked it the same question. And it said, yes, the spirit of the Third Amendment is that the government is prohibited from imposing any legal officials, not just soldiers but anyone from the government.

[00:03:55.79] So we had our first AI circuit split there. But that was meant to illustrate a larger point that the systems will give you interpretations-- and confident interpretations-- but in the background, there's stuff going on. There's different value judgments and assessments. And we are in a world where we need to talk about this, because the hypothesis is that certainly judges out there-- I don't know about them personally-- but you got to imagine in their chambers when the lights are dimmed, they're typing things into GPT-4 and asking, what's the proper interpretation of this contract term?

[00:04:31.94] Reasonable efforts or not, what's the ordinary meaning of this? And there are some benefits and some costs. So we're going to talk a little bit about this. But proponent of it said, this is a good thing. So what we have now is maybe a judge somewhere who has to interpret a term like reasonable efforts, but what does he know? Maybe he's flipping through a book, looking through some customs. They're kind of doing this ad-hoc arbitrary decisions.

[00:05:02.05] Maybe it's a new judge. Maybe they've seen-- but on the one hand, this is maybe arbitrary. Same thing from the constitution. So proponents of this said, yes, AI will allow for more consistent principled interpretations. Judges aren't exactly principles. On the other hand critics have a lot of things to say about this. That this notion of objectivity of interpretation by AI is illusory.

[00:05:32.92] So behind every interpretation whether the Third Amendment applies to only soldiers or the governor is a value choice. It's an interpretation and under the hood something's making a value choice, and we should be aware of these interpretations. And the critics also highlight the central role of humans in legal decision making. Like humans are important to make decisions about other humans, and that's a central principle of law.

[00:06:06.50] And others argue, wow, we've already had this debate before in plain old constitutional law before AI even came along. We have different interpretive philosophies. Some of you who are lawyers but others who are not, there are different interpretive methods. Things you might have heard words like originalism or textualism or structuralism or pragmatism, these are-- and we've had these debates over and over again, where some people says the text tells you everything you know, and others say no, you have to take into the context or-- and these debates don't settle anything.

[00:06:42.69] So this is where we are. We are in a moment where AI can read legal texts. And I want to emphasize like I did, they couldn't two years ago read it at the same level. So this is an interesting moment. We couldn't have had this conversation as reasonably as we did two years ago. So I'm going to ask the first question to Andrew Coan. As a constitutional law scholar, how do you think about the use of AI in the interpretation of the constitution? And is this a tool that will help ameliorate some of the seemingly intractable political or social disagreements about law in the US, or is this illusory?

[00:07:20.48] ANDREW COAN: Thanks very much, Harry. And thanks to all of you for coming to this panel and to everyone who's helped to organize this conference. It's really a pleasure to be here. I've spent quite a lot of time and quite a lot of my career thinking about constitutional interpretation, and a little bit of time thinking about artificial intelligence and large language models. So I'm going to offer some tentative thoughts here about the intersection of those two issues or questions.

[00:07:46.55] And really looking forward to hearing from my other co-panelists who I suspect are all quite a bit more technologically knowledgeable than I am. I want to raise just first a couple of questions, and then offer some tentative thoughts on this last question that Harry posed. Because I think when we talk about the use of AI-- and here I'm going to focus specifically on large language models in the use of constitutional interpretation.

[00:08:13.48] We need to be a little bit careful about what we are talking about and what we are imagining. There are a lot of different ways in which LLMs might be used in constitutional interpretation. Are we talking about the use of an LLM as a research assistant in a judge's chambers? Are we talking about an LLM that is being asked to predict

how a court is likely to rule by a lawyer or someone who has an interest in the outcome of a constitutional dispute?

[00:08:42.67] Are we talking about an LLM that a judge is asking to decide a case in some kind of global sense? Tell me the answer to this constitutional question. Are we talking about a more narrow use of an LLM to decide a case under a particular interpretive philosophy? Tell me how an originalist would decide this particular case? Or maybe even a judge using an LLM to ask or to answer a really narrow question, which is part of but not the whole answer or the whole of what we need to decide a particular case?

[00:09:19.23] I think all of these different uses probably have different pros and cons. And it's pretty important as we think about this question to disaggregate these various uses. Another I think important point to keep in mind as we think about these questions is that constitutional interpretation-- issues of constitutional interpretation arise in a lot of different institutional contexts. It's not only courts that confront constitutional questions. Lawyers and clients confront constitutional questions.

[00:09:49.10] Lawyers within the executive branch of the United States, for instance, and within various governmental departments at all levels of government encounter these questions, wrestle with these questions even in contexts were no dispute ends up in court. And of course, these questions arise at different levels even within the judicial system.

[00:10:08.34] The kinds of constitutional questions and the resources available to a district court confronting a constitutional dispute are different from the kinds of constitutional questions that are most commonly presented to the courts of appeals, which are different in turn from the kinds of constitutional questions which are most commonly presented to the Supreme Court.

[00:10:32.37] Among other things, the constitutional questions presented to the Supreme Court of the United States typically involve much harder issues. Issues on which the existing law provides fewer determinate answers and in which there is a wider range for plausible or reasonable legal disagreement. And I think the use of AI in these various contexts raises different questions, and it's important to keep those distinctions in mind.

[00:11:02.71] Finally, I'll just offer some tentative thoughts on the last question that Harry posed. Do LLMs or artificial intelligence offer the hope of achieving what is-- you can think of it as a kind of eternal-- although up to this point, unrealized dream by at least some thinkers in the realm of constitutional interpretation are providing, some kind of objective externally valid approach to resolving constitutional questions, that allows us to squeeze all moral judgment out of the process.

[00:11:38.85] And you can probably tell by my phrasing or framing of the question, that I'm pretty skeptical that this is likely to be the case. One of the reasons why I think I'm skeptical here is that of course, the choice to use an AI which is trained on a specific data set which embeds as many other panelists have already discussed here a number of predictable biases, is itself a kind of moral judgment or moral decision. It might be the right moral decision in some contexts, it might be the wrong moral decision, but it's a moral decision nevertheless.

[00:12:15.72] Maybe it's somewhat less obviously when you pose a constitutional question or a question of constitutional interpretation to an LLM. As when you frankly pose a constitutional question to a human interpreter, the way in which you frame that question has a lot to do with the answer that you are going to get. I spent a very pleasant Wednesday afternoon this week posing the verbatim questions presented in the Dobbs v Jackson Women's Health case and the Students for Fair Admission affirmative action case to Claude-3-Opus and to ChatGPT4, just as a small illustration of how this might work out in practice.

[00:13:05.88] I first asked the two models how they would decide these questions without giving any guidance about what interpretive philosophy they ought to provide or whether they ought to apply. Interestingly, both of them would have decided these cases in the opposite way that the US Supreme Court actually decided them. Dobbs was wrongly decided according to both ChatGPT and Claude Opus as was the Students for Fair Admissions case.

[00:13:33.33] I then asked them how they would decide these questions as of June 1, 2022. So based on the state of the law when these cases actually came before the US Supreme Court. Under the interpretive philosophy of Justices William Brennan and Thurgood Marshall or justices like them, I asked them how they would decide these cases as an originalist. And then I loaded the dice a little bit and asked them how they would decide these questions as an originalist but with a thicker, fuller more academically sophisticated explanation of what an originalist approach should entail.

[00:14:11.43] With a particular emphasis on one of the theoretical ideas to emerge from some of the most recent literature on constitutional originalism, which is the idea of linguistic indeterminacy, and basically that the constitutional text does not necessarily resolve-- the original public meaning of the constitutional text does not necessarily resolve all questions. Pushing the model to reckon with and grapple with the existence of indeterminacy in the original public meaning.

[00:14:38.49] And the need in some cases where the original public meaning of the text runs out, to supplement that meaning with additional considerations like judicial precedent or historical practice

or prudential reasoning. And by pushing the model in that direction, I was able to lead the model to reach the opposite results that it reached under a more straightforward and bare bones description of originalism.

[00:15:07.17] When I asked it to decide these cases as an originalist, it decided Dobbs the same way that the Supreme Court did. It decided Students for Fair Admissions in the same way that the Supreme Court did. And when I pushed it to grapple with the indeterminacy of the constitutional texts and the original public meaning, it flip flopped.

[00:15:24.12] Then I did a little bit of further experimenting where I tried to persuade the model that it had been wrong in the original answers that it gave me under each of these various interpretive approaches. And in each case, it was trivially easy to get the model to reverse itself based on a very basic one paragraph summary of the opposing arguments on the other side of the case.

[00:15:55.38] Harry told me after I engaged in this little exercise that there's a name for this phenomenon. It's called AI sycophancy, where the LLM tries to please the user to give you the answer that you're looking for. And I think that's actually a nice place to end.

[00:16:13.42] If the dream is that artificial intelligence and LLMs will provide some kind of objective basis for resolving our disputed questions of constitutional interpretation. And one that will allow us to squeeze all moral judgment or bias out of the system, the ease with which it's possible to lead the model to your own preferred outcome I think is at least a pretty strong cautionary note about that prospect given the state of the existing technology. So with that, I'm very happy--

[00:16:50.08] HARRY SURDEN: Thank you, Andrew. Let me just summarize a couple of themes in your terrific overview of AI. That was great. But I heard a couple of themes is one, despite the illusion, AI will not relieve us of the burdens of judgment. And indeed an interpretation done by us or the AI is still making explicit or implicit value choices about some issue in society. So that doesn't magically go away. So that's a great point.

[00:17:20.25] Another point I heard you make is the sensitivity of these models and the lack of repeatability. So it's well known that a lot of these models have intentional randomness built in. That's actually a feature, not a bug. It's designed to stimulate creativity. But this only enhances the idea of understanding the limits of these models. Because you could easily imagine a judge saying, I am a textualist. What is the textualist answer to this? Seeing an answer and being like, wow, OK here's the answer.

[00:17:52.41] Not realizing that the specific words they put in are leading it down a particular path. And they might get a very different answer if they left out the word textualist or they framed it differently. So this lack of repeatability. So making the legal profession aware of

that they're not getting objective answers, they're getting the illusion of strongly stated answers that are very sensitive to initial inputs.

[00:18:18.70] Let me open this up to the rest of the panel. Let's focus on the con law questions and we'll turn more to private law documents. Vivek, you look like you're raring to go.

[00:18:27.28] VIVEK KRISHNAMURTHY: Oh, no. I'm absorbing what you were saying. But on the point of repeatability-- and I guess this is for everyone to think about-- if we think about the legal system as a system, we do have-- and how cases move up the system to ultimately the Supreme Court, maybe that feature is a feature. In that we have different AIs and different runs of the model coming up with different answers. But it's not dissimilar to how trial court decisions you form a split that goes to the circuit et cetera, et cetera. So I don't know how we grapple with that.

[00:19:06.60] ANDREW COAN: So I think the difficulty of getting the AI to settle on a particular answer, getting different models or different LLMs to converge on a single answer could have some kind of utility in helping us think through the array of possibilities. But I think it underscores the extent to which the AIs are simply going to present us with new occasions for the exercise of moral judgment rather than allowing us to avoid the burdens of judgment as Harry put it.

[00:19:46.64] HARRY SURDEN: Megan, did you want to weigh in?

[00:19:48.09] MEAGAN MA: Yeah. So I think basically what I'm hearing from you Andrew is that right now, it seems a little bit unclear in terms of the user persona that's involved in the use of these tools. So when you talk about asking these questions and feeding it this judgment, like is it a law student that's kind of undergoing this exercise and trying to play out different scenarios? Or is it actually the judge who's focused on looking at this specific scenario?

[00:20:15.19] And why I find that interesting is because historically some of the predecessors to large language models, the transformers itself, we've seen there's a paper out by Dan Katz a while back called, Natural Language Processing in Law. And a lot of what they showcase is that pretty large language models, a lot of it was very, very task-centric down to they were doing exercises involved in information retrieval, on prediction. And all of these things that you've mentioned, they're rolled up now in a new feature known as natural language generation and where you're seeing a lot more of a dialogue.

[00:20:52.44] And I think the dialogue and the use of chatbots was the moment of spark, where suddenly it was much more interactive for the average person to be able to now reason with the US Constitution. And so to me I think the question is, as opposed to the back end me using this tool for to retrieve some information-- a one off kind of task-- because of this continuous iterative process, I might think that a use case that makes the most sense here is strategy and gameplay.

[00:21:26.82] It's kind of a transition now from just only doing tasks using these models to now thinking like, oh, I can work at a higher level of abstraction. And start to think if I'm using them and integrating my own judgments, I'm looking at what are my abstract goals here and trying to navigate that space. And I think that that is probably a distinguishing factor in why these models are so fascinating. Why there's all these prevalent risks is because it's no longer as specific as I want to retrieve one kernel of information on what this interpretation of this text means.

[00:22:02.97] Or like I want to look at all of these texts and find maybe a common pattern. What it is now is that there's constant back, and forth and constant iterative feedback that is going to change the way that I'm going to make judgments.

[00:22:14.89] HARRY SURDEN: This is a great point. And some of those themes echo the format which Professor Chris Goodman mentioned yesterday, the automation bias and automation difference. If something comes out of a computer, we see it as if it's mathematically correct somehow, where we might not give that same deference to a judge or somebody who might be expressing their opinion. I wanted to say, does this Vivek's comment resonate at all with your Silicon juries idea or?

[00:22:45.68] YONATHAN ARBEL: OK, so maybe a word of introduction. I dedicate myself over the last couple of years to thinking about how we can build AI tools for a better legal system. A more equitable, more access to justice. And so when I hear the problems that we're presented, I say, yes. These are problems. But my reaction is challenge accepted. Can we solve these problems? Yes, there is a true problem with sycophancy like the model will play along to your questions.

[00:23:16.59] There is a question of prompt engineer. Are you capable enough of articulating your question in the right manner? And what happens to the sensitivity of the model when you change a word? Will that produce a different answer? What about randomness? That's like a part of the thing. Can we find tools? Think about these problems proactively, and think about how can we build better tools. And the answer is, yeah, there are means of building better tools to address all of these questions, because there are more questions and even these questions have nuance.

[00:23:47.68] One thing we can do is listen, these things are not expensive. We don't have to ask them a single time and just satisfy ourselves with the first answer we get. We can ask them 1,000 times the same question. I mean, we can't do it to a human. That will be not nice. But we can do it to an LLM. Their patience will never run out. And we can collect the answers, and we can see if there are trends in the data.

[00:24:13.06] We can freeze the randomness by setting the temperature to zero, and then we can unfreeze the model and produce every time the same answer. And so all of these are problems that if we are sufficiently motivated, we see that we have an opportunity here to build better legal tools. We want to be motivated and take all of these questions, and respond to them by building better tools.

[00:24:38.26] HARRY SURDEN: So I'd love to follow up with that. So some of your work-- so we're moving now outside the constitutional law context to interpreting civil documents like contracts. And you've made the case that using AI to interpret contracts might provide some benefits. Can you say a little bit more about that?

[00:24:56.07] YONATHAN ARBEL: Yes, absolutely. So these tremendous models that are trained on vast amounts of data-- Llama 3 came out yesterday, $13 trillion tokens. That's just mind-blowingly large. It's more text than any of us will ever read in our lifetime. So how can we take these models and create interpretation machines that improve our current processes?

[00:25:21.50] Part of my work with Professor David Hoffman of the University of Pennsylvania, we have developed a notion of the idea of methodology of generative interpretation. Generative interpretation is exactly building the tools that think and are aware of these challenges, and we use them to probe those large language models and try to extract their understanding of language. So what do we do? We take a bunch of famous difficult cases in contracts mostly.

[00:25:51.08] We present them to the model and see how they reason about them, how they understand the language, questions of self interpretation. For example, we take the Katrina case. If you remember, Katrina left-- its sort of devastation-- wave of devastation in New Orleans. Tremendous suffering and tremendous loss of life and property damage. And then came the lawyers. And so the question became part of the litigation question was, can you sue the insurers for flood damage given that the insurance policy excludes flood?

[00:26:28.82] And this is the last thing I'll say about insurance policy because we're after lunch. And the question is that flood, because the argument was, the plaintiffs wanted to push forward the argument that it's not a flood if it's man-made. If it's a flood if it's rain damage, but not if the levees broke and that's negligence. Human negligence doesn't count. Like the word flood doesn't encompass natural causes. That was the argument. And it was a very creative argument. And the Fifth Circuit said, no, it's not going to work because a flood does cover all of these sources.

[00:27:06.33] And many people are very critical of the Fifth Circuit. They said, well listen, you're just reaffirming your business priors. So you're not really doing textualism. You're not truly trying to interpret.

What if we can take large language models that are impartial, that are reproducible, that we can very convenient to use. I know you alluded earlier to judges using them. We have surveys showing that judges actually experiment with them.

[00:27:32.94] What if we can take these tools and try to see how they understand the word flood? And what we show is that the court has actually got this one right, at least according to some definition of correctness. That the word flood as it is understood by large language models does encompass artificial and natural and man-made sources. And you can use it for many other cases.

[00:27:56.99] There is a Supreme Court case for example, Pulsipher came out very recently, where there was a question, can you read this document in more than one way? And the majority says, no there is no ambiguity. You can only read this statute in a single way. And you present that to the large language models, and you see that the way they understand language is actually quite ambiguous.

[00:28:19.05] We also see in surveys when we survey people, but it's very difficult to survey people for every question of interpretation. We can do it with LLMs. And in both cases, we see this sort of ambiguity emerging. And that's a very powerful tool we think for judges. When they are very certain about the way they read language, it's a check to say, well, maybe this language is open to multiple interpretations.

[00:28:43.80] And I want to expand also the view of who's exactly interpreting legal documents? I mean, we probably typically think about judges in their chambers trying to picking dictionaries and trying to understand language. But legal interpretation is a task that we all engage in. Regulators engage in, citizens engage in when we're trying to abide by the law, we're trying to perform our contracts. This is all involves a measure of interpretation.

[00:29:12.15] And we have a problem there in contracts which is my main area, that people don't read contracts. They're too long. People tell me they're boring. I take offense to that. [LAUGHTER] And what we've done, we experimented with hey, can we use LLMs as smart traders, like your smartphone? And what we show is that-- what we found is that you use LLMs you take all those long terms of service-- all the iTunes privacy policies-- you cut them down by almost 60%.

[00:29:46.51] You shorten reading time by 14 minutes and 49 seconds on average. That's quite significant. You reduce the level of the text from a college level to the level of a fifth grader. And guess what? You're not missing any critical legal information as evaluated by experts. So we see tons of potential.

[00:30:09.77] And when I'm so motivated to think about your problems as challenges rather than fundamental problems not to think-- and I'm not saying, this is how you put it. This is what motivates me. I think we

have a great tool, and we can build stuff, and it's not a project for any single person. So I invite all of you all to think with me about that.

[00:30:28.96] HARRY SURDEN: These are some great points. Thank you Yonathan. I want to pass this off to the panel in a minute. But I'll just summarize some of the themes. So one theme is you might have strikingly different considerations in the private contract context than you do in the public constitutional law context. So we might be more willing to engage in contractual interpretation by AI versus constitutional interpretation, which might affect the entire country rather than individual people.

[00:31:00.50] Another thing I heard is we shouldn't automatically assume that judges are error free and excellent. So who knows what the Fifth Circuit has their own opinions about things these days. And in particular, if they interpret the word flood, who's to say that their interpretation is rigorous or transparent or not motivated. On the other hand, I was reminded of the idea where you said that the LLM agreed with the Fifth Circuit whether you did experiments to test for the sensitivity of the input, and whether if you rearrange the prompt slightly it might come out with different.

[00:31:39.41] But your point is well taken that that is at least a useful tool. And then third, I heard this idea that it might educate judges who maybe have narrow focus that there might be more than one way to interpret things by putting it through the LLM that can show you there's ambiguity you didn't see. So let me pass these questions out to the panel there.

[00:32:05.61] ANDREW COAN: I can start. Thanks Yonathan. This is really interesting. And I think you've prompted me to want to clarify something about my own remarks, which is just that none of the issues that I was discussing at the end of my remarks is one that I see as necessarily an insurmountable problem. Maybe not even a problem at all, it just seems to me that we should be clear about-- given at least the current state of the technology-- the extent to which moral judgments of the user of AI in any particular context are likely to play a role in the outputs that we're getting.

[00:32:42.64] So I'm curious-- a couple of questions-- whether you see the large language models in the context of contract interpretation as a route around moral judgment by humans, or simply as a really powerful auxiliary to moral judgment by humans? And second-- just to follow up on Harry's question or observation-- I'm curious how much testing of the sensitivity you did? In particular, how much you attempted to talk the AIs out of their original answer.

[00:33:20.81] Because it seems based on my playing around, I think your study was much more systematic than my little noodling around on Wednesday afternoon. But my intuition based on my own experimentation with the models at this point is that any reasonably

competent lawyer can pretty easily talk them into a different result by emphasizing, for instance, the context in which the word flood is used and the assumptions that people reading an insurance contract might have about what would be covered and what would not be covered.

[00:33:55.11] And the role that under modern linguistics or the philosophy of language we understand context to play in linguistic communication and the formation of ordinary meaning of a document and so on and so forth.

[00:34:11.61] YONATHAN ARBEL: Yes. So no surprise that I actually agree with the motivation behind your questions. Like hey, how far can I push it? And what happens when I push it? And that's such an important question for the legal process. We want to make sure that the outcomes of legal cases reflect the merit of the case and not so much the identity of the parties, or maybe you have a super powerful attorney who can frame things so ever so powerfully.

[00:34:38.48] And well the truth of the matter, part of what we're teaching here our law students exactly that how to prompt engineer humans. How to frame your case such that it will be compelling. And that's maybe also sets the ground for a point that I think gets lost very, very frequently, which is the compare to what-- and this is a normative judgment. But I don't think our legal system right now is the best anyone can ever aspire to in terms of truth searching or vindicating justice.

[00:35:17.55] I think there are many structural problems we have today. And when we are critical of the models, we should also say it's not good, but is it better? Not rather whether it's the best option we have. Now, well one more point about that. We want to make sure that when we ask a model question, we get an answer that's not very sensitive to the prompt. And we can do it. We tested it. And we imagine a world where both parties will present many different prompts to the court and the court can say, well, here are 1,000 different prompts we're checking.

[00:35:53.40] And we're seeing what is the correct interpretation of flood, or what is the major trend within all these different prompts? One of the beautiful things we can do that. That's something that actually can be done. We cannot do the same with jury instructions. We send a question to the jury, we frame it, we use words. These words decide the case many of time. You can imagine different jurisdictions, and in fact, this is part of litigation. Fighting over what goes out to the jury.

[00:36:23.34] We have one chance. We have freedom when it comes to LLMs to experiment more. Now in terms of moral judgments, generative interpretation as we think about this method is one half of the interpretation project. Interpretation is not just deciding what words mean, it's also what they should mean. And here we imagine

that this is where the judge steps in. This is not about replacing the moral judgment. But I think in your comments in your questions there's like, well when we decide language there is some normative element involved.

[00:36:59.25] You can really prune the normativity or the ethical judgment from deciding the meaning of words. And here I have one reassuring finding from recent literature. The literature on silicon sampling or I work on silicon juries. It turns out people have started asking, hey, instead of surveys in the social sciences, can we ask LLMs to be our respondents? All these polling questions we read, can we use LLMs instead of flesh and bone humans?

[00:37:37.30] And before we do that, we want to make sure that the responses actually correspond to what humans think. And the findings so far is that LLMs are actually kind of good in getting broad dispositions, like what Democrats think about Republicans. They can capture it quite well. They cannot capture well any sort of nuance. This is what would a Democrat from Chicago would think about a Republican from say Alabama.

[00:38:05.80] But yeah, they don't do well there. But there is-- they do encode and embed moral judgments into these models that correspond either through the training data or RLHF to actual human judgments. So there is some of reason to be optimistic.

[00:38:26.79] HARRY SURDEN: Yeah, that's a great point. And it does echo some of the themes we heard yesterday of if we're going to do something like that, we'd want to be concerned that the training data is representative of different demographic groups where it is at least historically been skewed and not representative. Let me pass this out. Megan, did you have a comment on the question originally directed at Yonathan about the benefits?

[00:38:53.51] MEAGAN MA: Yes. So I know Yonathan doesn't want to talk about insurance but I guess I will. We have an insurance initiative over at the Stanford Center for Legal Informatics or Codex. And one of the current tools that we've been working on with re-insurers is actually this notion of elasticity. So it's basically how and why we stretch our words. So not in the economic sense. And really it's supposed to be a play on vagueness. When and why do we use vagueness?

[00:39:22.23] And can we detect actually moments of intentional vagueness versus unintentional vagueness? And the idea is that unintentional vagueness is actually costly especially for reinsurance, which for folks that don't know, it's actually insurance for insurers. And so the risks are really at stake here. And one of the things that they frequently talk about is this notion of silent risk. Did I craft my policy such that I inadvertently wrote in or am covering some sort of risk that I don't want to?

[00:39:51.50] And so we built this tool that essentially breaks down the clause and identifies all these moments of high elasticity. And then we allow these underwriters-- so the contract drafters-- to actually say, yes. This is a moment of elasticity that makes sense given the fact that we value these long-term relationships and these highly negotiated contracts. Or this is like business operations as usual. This is just the typical language.

[00:40:17.80] And essentially these models are really great at giving almost like a write up and a quick report to illuminate an alternative perspective as to whether or not the underwriter actually was vague in the way that they want it to be. And it gave a better lens, a secondary pair of machine eye, so to speak. And I think that that's really pointing to Yonathan's point of like, with these large language models, it's not just kind of enabling interpretation, it's also actually helping the folks that are in the drafting position.

[00:40:49.78] That it's more than one user at play, and it's really about the use case that we're working with here. So even if flood is-- we always think insurers are the bad guys. But if we offer tools like this, maybe they have a little bit more clarity as to how they're writing these contracts.

[00:41:06.92] HARRY SURDEN: That's a terrific point, Megan. So those who are in the law know that ambiguity and vagueness is not always a bug. It's sometimes a feature. So it's thought of outside the law why do you have a word like reasonable? Something like that, but within the law, often that gives flexibility and judgment and discretion later in time, and that can be beneficial. On the other hand, sometimes vagueness actually and ambiguity is a bug especially if you didn't mean it to be there.

[00:41:36.65] So I love this idea of using AI as almost like a vagueness spell checker in a sense that helps us. Do others want to weigh on this idea about using AI in this context and intentional or unintentional vagueness? Any thoughts? All right, so Vivek, so one of the themes that has come up a bunch is that we framed it as this is a judge late at night when nobody's around using GPT-4 secretly. But that's not really the way it is.

[00:42:18.58] One of the real benefits is that lots of different users can use GPT to maybe understand the law and make it more accessible. So I want to throw it out to you, Vivek. Tell us about the different contexts in the law in which AI large language models can be used and some of the benefits and risks of that.

[00:42:37.88] VIVEK KRISHNAMURTHY: Yeah, absolutely. Thank you Harry. Thank you Silicon Flatirons. Thank you all of you for being awake after that great lunch. So let me begin with a question. Who here is a lawyer or has legal training? So about half to 60% of you. Congratulations, you represent 0.4% of the American population. And

we are the country that has by far the highest rate per capita of lawyers in our society.

[00:43:07.06] So we are really not following Shakespeare's advice on killing all the lawyers. OK, so here's what I'd like to start. It was tax day this week. I hope you all filed your taxes. And Donald Rumsfeld, the former Secretary of Defense, every year-- may he rest in peace-- would send this letter to the IRS saying, here are my tax returns. I have no idea if they are correct. I'm a college graduate. And I can't tell because our tax code is so complicated. That's terrible.

[00:43:35.04] It reflects a travesty that a well-intentioned person can't tell if they are filing their taxes. I'm a law professor, and I also filed my taxes. I have no idea if they're accurate or not. But I did rely on a tool. I relied on TurboTax which is an earlier instantiation of AI. So go back to Harry's point this morning about the different schools and the history. It's an expert system as far as I can tell.

[00:43:58.93] Someone went through and programmed at least some of the tax code into TurboTax. And there's points where you hit it and say, I can't help you with this foreign tax credit. Go talk to a human at this point. But we've encoded the law in a primitive way. In a way that actually produces very high quality results that compare very favorably to the best accountants or tax lawyers in the world.

[00:44:24.35] So I'd like to make that my starting point about how we think about AI and the law, which is that-- and this builds on Yonathan's point-- which is that everyone in society makes legal interpretations and legal decisions and has to follow the law. And the law goes well beyond the Constitution and the statutes and the rules. It includes building codes. When I'm deciding, can I do this electrical work by myself at home? I probably shouldn't. But there's law around that.

[00:44:52.54] And to the average person who is not an expert in a particular area of law, not only is the law unknowable, just the fact whether there is regulation that applies to that question is often very difficult to discern. So I think we've sort focused a lot on, again, what Rumsfeld would call the known unknowns. The hard questions. How should the Supreme Court decide Dobbs? How might it do that? And those are a very, very, very, very small percentage of legal questions that arise in society.

[00:45:23.23] They are the very top of the pyramid. Even things that make it to a court, I would argue, represent a long tail on the distribution of legal questions that come up in society. So to me the fact that all of us in a society of rules-- and there are so many rules-- have to follow these rules and interpret them creates a tremendous opportunity for better legal tools. Tools to uncover the fact that this is a regulated area, and tools to give us first approximation answers to questions that are probably very easy for experts.

[00:45:56.68] So my tax situation in the realm of tax situations is probably pretty simple. I don't have anything crazy but it was hard enough for me. So we think about that across many different fields, fields of technical regulation. We think about frontline government officials that have to make decisions that sometimes implicate constitutional values. Can I stop this car? Is this a search incident to arrest?

[00:46:23.44] Those are all questions for which arguably there's actually quite a bit of settled law, and that 99 trial courts out of 100 would probably resolve in the same way. But again, to a person without legal training or expertise in a particular area of law, those simply can't be resolved. So I think thinking about the application of AI in outside of contentious cases, thinking about compliance.

[00:46:52.78] And a lot of what lawyers do in practice and what people who are in regulated fields do, is attempt to comply with different kinds of rules. Some of them have legal force, some of them might be professional standards codes of ethics, you name it. Engineers are regulated profession as our doctors as our dentists as are everyone else. So I think that's an area where we really should think. And I think it's an area where it's perhaps less contentious.

[00:47:18.13] Where we could actually look already at a system like GPT-4 and find probably convergence on the answer to these easier questions, because they are easy and they are more settled but they're still unknowable and undiscoverable to folks without that expertise. So the last thing that I wanted to point to is that I do wonder if there is an opportunity beyond the Q&A function of something like GPT, to either surface rules or to help provide some, let's say, easy legal advice to rethink how it is that we apply law in society?

[00:47:57.50] So I'm going back to my tax example where we have a form, an extremely complex one. I'm a naturalized American citizen and let me just tell you, that was a hell of a thing to navigate. And I look at people who have no legal knowledge and perhaps limited English proficiency and say, man, how the heck did you navigate that system to get your green card or whatever benefit it is?

[00:48:21.74] So obviously there's a lot of innovation happening in terms of access to justice, in terms of building better forms, building better legal tools to allow people to access benefits. But to me, I wonder if we move our focus away from legal decision making to a compliance orientation of helping people follow rules or seek benefits to which they're allowed to and leverage these technologies if we may get more bang for the buck.

[00:48:51.44] If we could really improve the quality of Justice and of access to justice in our society, and perhaps avoid some of the harder legal questions that ultimately we are going to want to have humans decide for those moral reasons that I think we've been surfacing in this

panel because they do involve hard moral calculations. But again, filling my taxes out correctly should not be that kind of decision.

[00:49:15.26] HARRY SURDEN: Great feedback. That was terrific. So just to sum up a couple of things, maybe we would imagine if Rumsfeld were still around today, this year he'd be writing the IRS, this year I finally understand the tax law thanks to GPT-4. Thank you. And another theme you brought up is you make a great point. So the scholars from the science and technology studies systems and other allied fields have been writing for years that socio-technological systems embed values.

[00:49:47.34] So something like TurboTax, they read the tax code, implemented it in some system, but those are value choices that we just talked about. Not unlike saying that the governor is a soldier because a soldier really is a proxy for government officials. So if they list something as a business expense or not, they're making a decision. So value choices get embedded one way or another, and there may not be any way to avoid them.

[00:50:14.68] And then the final point I really liked is, in law school, of course, we focus too much on judges and we shouldn't. But we're all citizens governed by the very laws around us. So this is the idea of the law of the ground, law on the ground, and a lot of these day-to-day decisions may be way more impactful than the output of GPT-4 on one judge's interpretation. Let me throw the same question out to the rest of the panel.

[00:50:43.17] Yonathan, you had echoed some themes about different user like outside of judges, different contexts. I saw some other thoughts out there.

[00:50:52.68] YONATHAN ARBEL: Not just echo. If I could, I would say the same thing. Because I think that's absolutely right. I think the barriers to access to justice are huge and some of them are artificial. And I have a question for you. We have-- in the context of unauthorized practice of law-- we have decided as a society that getting legal advice from someone who does not have our level of experience, does not have our license, that is impermissible because they might get the wrong advice.

[00:51:27.25] You make the point that there are so many decisions in our lives where the stakes are not that high, the cost of mistake is not very high. And so even the rare case where GPT-4 might hallucinate or at least maybe this is I don't know. How I'm understanding you is we're still like that's better than the status quo where people just don't know what the law is. But I can imagine a near future where there will be a push-- a strong push-- to make legal advice via LLMs illegal.

[00:52:00.42] Unauthorized practice of law and there will be cases too-- I'm sure there will be cases of people who've lost significant economic rights, because they've relied on bad advice from GPT-4.

How do you think we as a society should decide on whether LLMs should be given the permission to even generate legal guidance?

[00:52:22.51] VIVEK KRISHNAMURTHY: So I think there's a few different ways of thinking about this. First of all, I will say that I think the unauthorized practice rules are extremely restrictive, and that admission to the legal profession is too restricted in many ways. And that again, I am completely incompetent to give you family law advice, despite the fact that I'm a member in good standing of the bar of the state and of several others. So don't come to me with those questions.

[00:52:47.97] So the UPL rules are not-- first of all-- actually an effective check on competence. We use other mechanisms to govern the competence of lawyers. Which is that even if you're duly barred, you can you actually ethically give this advice. But there is a distinction that we've long drawn between legal information and legal advice. So I think where it might get interesting is to think about feeding into GPT a series of facts and saying, give me advice on this facts as opposed to the issue spotting.

[00:53:21.41] And even if we were comfortable with a regulatory framework whereby we could provide issue spotting via automated means, I think that would be a win to say that, actually you might be violating some laws here. There might be some relevant regulations. Here's what they are. Go talk to an expert on this. So to me-- I just want to build Harry's point. I think about systems, socio-technical systems and the law as a system and all of its functions-- the regulatory functions, the way that we furnish advice-- nothing in the status quo necessarily needs to be that way.

[00:54:03.02] And other societies do have different ways of provisioning legal advice. So in civil law societies, notaries do a lot of stuff. They're less regulated. It's easier to become a notary. And they can deal with entire classes of legal issues much more quickly and efficiently than a lawyer would. A lawyer would deal with something contentious. A notary would draft your will and draft a contract and do this and do that.

[00:54:30.47] So I do think that as this-- it's coming. The systems are going to generate this advice and these answers. So I think it's going to put pressure on the legal profession. That's probably a good thing because the system we have is not that great.

[00:54:44.36] HARRY SURDEN: Anyone else? Andy, did you want to weigh in?

[00:54:46.22] ANDREW COAN: Yes. Thanks, Vivek. That was really great and thought provoking. I wonder if we might-- from the discussion as a whole and in particular from your presentation-- distill and whether you would agree with this distillation of a couple of principles where for thinking about when the value added of LLMs might be greatest at least as compared to the alternatives.

[00:55:11.84] And two points that stand out to me from your presentation are, that in contexts where there are relatively clear answers, the value of LLMs versus more costly alternatives-- and almost every alternative is going to be more costly-- might be quite high.

[00:55:27.80] And second, in contexts where the stakes are relatively low and where the number of impacted persons is relatively small, this too might be an area where LLMs have greater value added compared to human alternatives or some combination of human legal expertise and LLMs. And I wonder if you would endorse those principles and whether there are other principles that you think might emerge from your account?

[00:55:52.13] VIVEK KRISHNAMURTHY: Yeah, so I think that's exactly right. I think there are some areas of the law where answers are relatively clear and the stakes are relatively lower in terms of the affected parties or the impact of something. And I would just look at where we have let's say low cost sources of legal advice. So look at forms, standard forms. Think about the Nolo series of books. How to write your will.

[00:56:17.62] Are those areas of the law where people need legal advice to structure their affairs or some clarity about what they're doing in which automation could really help? And I'm not saying that necessarily the models of GPT that we have right now that are the OpenAI GPT-4 as opposed to the versions that are made specifically for the law with more safe safeguards and/or other approaches to developing these systems. I think I'm open to all of those possibilities.

[00:56:49.72] Like what if we did use an LLM to help us develop better expert systems? That could be an interesting approach to say that, here is a system that we've developed for family law problems where we've used LLM to identify a range of issues that come up in large corpora of text. And then we devise a tool that helps people navigate the system. So that could be an approach to improving access to justice within that area of lower stakes, lower impact issues that would have considerable benefits in my view.

[00:57:25.13] HARRY SURDEN: Megan, did you want to [INAUDIBLE]?

[00:57:27.02] MEAGAN MA: Yeah, so I think one of the interesting themes that sort of surfaced is how well actually do these models perform relative to lawyers? And if we're saying that lawyers are the beacon behind who we should seek legal advice from, can they really have the credentials to stand behind that? And one of the efforts that actually we're working on together, Harry, is this bench marking effort.

[00:57:53.06] And a lot of the literature out there right now-- you've seen all the flashy headlines like GPT-4 passes the bar. ChatGPT goes to school, all of those things. But what's been missing from these empirical studies is that it has not clearly actually performed a legal

task that lawyers do on a day-to-day basis. We don't actually then have a realistic metric between what is good legal work versus not.

[00:58:16.71] And so one of the things that we're trying to work on here is we're doing a blind evaluation, and we've partnered with DLA Piper on this where we'll have a set of mid tier-- not mid tier-- but like mid-level attorneys, three to five years. And they're given a legal task that they do on a day-to-day basis. We give that task as well to GPT-4 or whatever is set of state of the art models. And then we ask senior partners to blind evaluate to see what their preference is.

[00:58:43.19] And I think a lot of right now what we evaluate these models on is purely on this dimension of accuracy. But for many lawyers, accuracy actually is the synthesis of the information and then driving a good argument behind that. It's not accuracy from the binary of a yes/no answer. And so doing this type of benchmark actually could probably anchor us better in how far off are we actually from these models doing legal work and performing our legal tasks, and where can they actually work together with us as kind of our handshake of human-machine colleagues?

[00:59:18.41] And so that's a little bit about how I'm seeing it is, at the moment right now when we're talking about these simple tasks, these easy tasks and being able to outsource certain legal activities to these models, we should first have a better metric around where are those easy tasks located. And Harry's done a great job in identifying all of those and doing those types of experiments. But more empirical work in this space could certainly illuminate where there is necessarily this breakdown in the argument of UPL for example.

[00:59:49.94] HARRY SURDEN: Yeah, thanks Megan. That's a great point. And I just want to echo one of the points you made, which is we're very much concerned that AI models give good legal advice if they're being used by people. But also lawyers make mistakes. And only in the most egregious instances do we know when lawyers mess up when there is either a bar complaint or a malpractice lawsuit. But in the run of the mill case when a client goes in, talks to a lawyer and gets bad advice in an office, we don't see that.

[01:00:22.89] So we want to make sure that we some-- we want the AI models to be providing solid advice. But we don't want to hold them up necessarily to a much higher standard than lawyers where we don't really have many benchmarks on how good lawyers are, generally speaking. And then to just echo your points, one of the themes I've heard is we all have to comply with the law whether or not we have lawyers. So something like 80% of Americans can't afford a lawyer for a civil matter that they have like immigration or family law or what have you.

[01:01:00.18] So their alternative is to either guess, ask a friend, Google, maybe ask a trusted community member. So wouldn't it

potentially be better if they're getting-- I call this the advice absent baseline. They're at least getting something from an AI large language models in when the alternative is just taking a guess in a complicated world. Let me throw out the final question then we'll open it up to the audience for question.

[01:01:29.44] So looking forward, and remember we're skeptical of predictions but we'll do it anyway. Where do we see AI being usefully integrated into legal interpretation in the next three plus years? And then how can we ensure that AI-driven legal interpretations are transparent, explainable, accountable and equitable especially when dealing with crucial legal documents? So this is a free for all for the panel.

[01:02:01.61] MEAGAN MA: I might start. So I think if I look at some of the analog practices around interpretation, so I come from a linguistic background. And we pull in linguistics. And since it's after lunch, I'll use a quirky food related example. So there's a case that was presented before the Massachusetts court a while back, where the court actually had to decide whether or not a burrito was a sandwich.

[01:02:26.21] And so they decided to use actually linguistic theory on it. And specifically within semantics, they decided between componential theory versus prototype theory. And just a brief sort of overview on that, componential theory is that it's the sum of the parts is what it means. And the latter on prototype theory is, let's imagine in our mental models what's the classic example. So prototype theory, a great example is, when you think about bird, you think about a robin, you don't think about a penguin, for example.

[01:02:56.22] And so the court was trying to negotiate a burrito is not in the prototype and so it can't be a sandwich. And other arguments were like, well a sandwich involves some lettuce and whatever, and a pita bread, is that actually like the bread that we want? And so then it became infinitely regressive at that point. And so that's why they defer to prototype theory. But the implications of that case actually was that really it was an exclusive rights case.

[01:03:27.99] It was a breach of contract case. And the background to that is Cordoba, I think they sell like Mexican food. Vaguely it's a fast food restaurant. They were imposing on Panera's Sandwich King territory basically. And they had an exclusive contract with this food court saying that I am the sole seller of sandwiches. And Cordoba like be on your way because you're selling sandwiches.

[01:03:54.36] And so ultimately the court was trying to decide whether or not Cordoba's presence in the food court actually breaches this exclusive right. And so to me, this is kind of an interesting case, because what we're trying to draw here is there's obviously things that are outside just this linguistic semantic theory of whether it be prototype or componential theory.

[01:04:16.08] And I think a lot of what we've been thinking about with AI is actually a focal point on context. And what's been fascinating is that a lot of older linguistic techniques that have been used in court is again, drawing from textualist or modern textualist theory and textualist methods. And we've seen, for example, the use of dictionaries deference to external sources, and trying to draw on dictionaries as some sort of neutral source.

[01:04:43.30] And the other sort of newer linguistic technique that's been out there is legal corpus linguistics. And what's fascinating about legal corpus linguistics is it's very, very old school NLP. Such that you're choosing a corpus-- choosing just a corpus of text for example-- and then deciding how it's being used and then the words that are nearing the use of a particular word.

[01:05:08.08] And then so we've seen a number of cases actually. In the most recent case I think we've noticed that actually at the end of the day, the courts really are choosing linguistic theory because it has sufficient flexibility. You've seen a lot of courts go dictionary shopping for example. With dictionaries start to surface kind of back in 1893, but the use of dictionaries actually-- deciding say between whether a tomato was a fruit or a vegetable back in 1893-- was based on tariffs.

[01:05:38.44] Whether or not you should tax a tomato as a fruit or a vegetable, and it has very, very different pricing as a result. And so why I mention all of this is that to me it's almost like a natural progression into our current use of large language models. And going back to earlier Andrew's example of context, what do I actually put in the context window or when I'm going through prompt engineering? That's that level of flexibility that's existed when judges have gone dictionary shopping or when judges have chosen the right corpus. It's about that sort of manipulation.

[01:06:14.47] And so in this era now of using large language models and context is kind of the key component, that's where we've actually learned to become a little bit more intentional. In the past like with these kind of judgments and deference to dictionaries or legal corpus linguistics, it almost seemed like you're kind of choosing an impartial source. And that's what the judges wanted to present in this case.

[01:06:37.63] But with large language models, we've actually become a little bit more cautious so that the intentionality of what goes into that context window, we're already sensitive to. And I think to me that might be potentially a benefit of being able to use large language models in this kind of space, because suddenly we've become aware of the role of context and in the use of language as opposed to just what does it mean? What does it not mean? It's actually how is this word being used and how is context being integrated?

[01:07:10.12] HARRY SURDEN: Megan, those are some terrific points. And I like that you reminded us that the law searching for external

authorities that purport to relieve judges of the burden of judgment and mask policy decisions has been going on for centuries. Andy, did you want to make a point?

[01:07:30.03] ANDREW COAN: No. I just would echo what Harry said. Those are excellent observations. Megan, I think very perceptive. And I guess I would just flag the extent to which they underscore the point that I was really trying to make at the outset, which is that we are not by using LLMs-- either in conjunction with or as a substitute for human judgment-- squeezing these moral judgments out of our legal work. We are in fact, making them in a different way.

[01:08:10.23] That's not to say that using LLMs-- again, either as a substitute or as a complement to human legal analysis-- is better or worse. It's just to remind us that the moral judgment is still happening, and that it seems important to surface that to make sure that that aspect of what's going on is transparent, and that LLMs not function-- in the way the dictionaries I think often have-- as a kind of mask that prevents the public or other stakeholders from understanding the extent to which moral judgment is still being exercised.

[01:08:53.37] HARRY SURDEN: Excellent point. All right, so I want to open this up to the audience. And our tradition is to have a student ask the first question. Do we have a student volunteer, a current student? I see some students in the back there. That are very-- all right, thank you. Thank you. And then we'll turn it over to the audience, but this is the Phil Weiser, Paul Ohm rule.

[01:09:24.89] AUDIENCE: Really interesting panel. The DLA Piper experiment that the Codex group is running, that sounds super cool, really interesting. I guess my question is related to bias in all of these conversations. And so do you think that bias is easier to correct in AI algorithms as opposed to human judges? And if so, what does that mean for using AI to render judicial decisions or how judges should use them?

[01:09:54.58] HARRY SURDEN: That's a great-- I mean, I have some many initial thoughts which-- so right now large language models are what are known as not interpretable. So we can't really look at one trillion numbers and figure out why it produced the word Paris instead of cat. But I have hope in the not maybe in the next five years based upon research, we will have more of an examination and interpretability to understand why they produce that.

[01:10:19.99] Whereas it's probably unlikely we will look inside a judge's brain to figure out why she came to one decision or another. But I think right now they're both pretty uninterpretable.

[01:10:35.52] YONATHAN ARBEL: So maybe to add something about the bias problem which is a very real problem. And it's a real problem that not because the model is not good, but because the model reflects the underlying data. And of course, when you read the entire

internet, you're going to find a little bit of bias. But of course, the beauty of that-- and this is part of what fuels my optimism-- is that we can actually do things about it.

[01:11:00.43] And this is the-- I hope maybe the builder mindset maybe this panel shares. So what can you do? So we have the problem. We don't know in the black box the activations exactly like whether there is bias. But we can test by having many runs and changing just the identity of the plaintiff, does that change anything in the judgment of the model? We can do other weird things. Apparently you can take the model and add to it mechanically, be cheerful, and then the answers become more cheerful.

[01:11:39.13] Be sad, and then the answers become-- we can add I don't know like surgically how that would work with humans, but we can throw in a component and tell it, hey bias-- bias maybe has two meanings here-- bias yourself in a way that is actually like pro-social anti-discriminatory. We can do a lot of things that we cannot do with humans. We cannot-- look, human judges are black boxes as well. We don't know whether they're biased or whether this is like their actual considered judgment. And we have opportunities here. This is why I'm optimistic.

[01:12:19.39] HARRY SURDEN: That's a really good point actually. It's like a systems engineering approach. We can't guarantee any particular output, but over time we can say, here's the likelihood of within this range of getting certain kinds of outputs that may or may not be biased. Really terrific point. Any other comments? Another question, this gentleman has been waiting patiently.

[01:12:45.68] AUDIENCE: Thank you, very enlightening. I have what seems to be a really simple question and maybe it's not, but it seems like half of Coloradans are tied into an HOA of one sort or another. And/or one third of the country is tied into an HOA, so it's kind of a pretty huge deal as far as contract law is concerned. And we had a new old law, House Bill 22-1137 that Governor signed, Polis.

[01:13:13.46] My question is, it seems to be a real gray area on community managers not being licensed and the enforcement of covenants. And there's a lot of wiggle room. There's a lot of authoritarianism. And I was just wondering, the big push is now not the collection agencies that are making-- had a real feeding frenzy and all this litigation or against the New House Bill, but it's for the people as it were. It's in their best interest legally.

[01:13:43.82] But we're involved in still litigation and it's thousands and thousands of dollars now. And my theme not being a lawyer is just ethics. So all I could do is go to legal theory here at the law school and just do the legal theory. So what's ethical, what's fair gets into estoppel and all that. So my question to Mr Arbel or Mr Krishnamurthy might be,

in this kind of gray area that affects millions of Americans, it's pretty foundational.

[01:14:13.92] What might you-- have a spin on that, should these property management companies, should they be tied into AI to make it more equitable, more fair? Because the licensing is wishy-washy. So it's complicated but it would seem to be simple. So thank you very much.

[01:14:32.97] VIVEK KRISHNAMURTHY: So let me just abstract from your question something larger. Which is that decisions get made about us a lot by lots of different people in contexts where there's different degrees of regulation. Some are highly regulated, and the rules are very clear and determinate, and then there's a situation that you're experiencing of the HOA is where things-- I mean HOA has a lot of power over you.

[01:15:02.19] They can Levy some huge assessments and all this stuff. So to me I think it's really interesting to think about, can we improve the quality, predictability, accuracy and fairness of those kinds of decisions using various kinds of automation? And I would say-- I don't know if we can do that for sure with GPT-4 as it is today. But I'm optimistic that we can. And I think that we should.

[01:15:30.18] There's a lot of work that's done on a different kind of private decision making and automation of content moderation. That's obviously something where a lot of us in this room have thought about or written about, where we do actually feed in something that looks like constitutional values into automated systems to decide does this come out, stay down or stay up? And I'm not going to take the time to move to the next panel away. But yeah, I would say that it is an area that we should think about.

[01:15:55.84] HARRY SURDEN: 30 second. We're almost out of time, but second reactions to the question here.

[01:16:00.13] YONATHAN ARBEL: 30 seconds. We have a model of how people come to assert their rights, the name, claim, blame model. And the problem we have right now is a lot of people don't to name the problems that they face. When the HOA discriminates against someone on the basis of maybe racial prejudice or something else, people don't always know that they have experienced a legal wrong, which prevents all the downstream ability to claim your rights and fight against that.

[01:16:28.39] And I see huge potential for people just, hey ChatGPT, what do you think? Is that OK for them to tell me that I can't be outside or can't have activities? And that that'll be I think foundational and I think intersects with your view of use case based.

[01:16:49.38] HARRY SURDEN: Terrific point. So unfortunately we are out of time, but please join me in thanking this terrific panel here.

[01:16:54.52] [APPLAUSE]

AI Speech and the First Amendment

https://youtu.be/7HyVxlYUsGM

[00:00:01.97] CATHERINE FERRI: I'm going to introduce our final panel. My name is Catherine Ferri, and I'm a third-year law student here at Colorado Law. I have the privilege of introducing our final panel, AI speech in the First Amendment.

[00:00:14.27] So the First Amendment has long protected the speech rights of human, and corporate speakers, and listeners in the United States, but rapid advancements in artificial intelligence have led to a new era, in which AI-generated speech can match or even surpass human speech in terms of sophistication and substance.

[00:00:31.49] This development has raised complex questions about the intersection of AI and speech. How does the First Amendment apply to speech generated by machines, if at all?

[00:00:40.11] Do individuals have a right to access AI-generated speech in the same way they do with human-generated content? What are the concerns about potential censorship and the role of the government in the realm of AI speech?

[00:00:51.26] This panel will explore the various implications of AI-generated speech on the First Amendment. Panelists will discuss whether the right to free speech and the right to information extends to the creation and dissemination of AI-generated content. And if so, to what extent.

[00:01:05.86] The panel will also explore the potential differences between AI and human speech, as well as the challenges it creates with speech that is potentially defamatory, misleading, or inaccurate. Ultimately, this panel aims to parse what the First Amendment future looks like in this new world of AI.

[00:01:20.67] Moderating this panel is Professor Blake Reid. Professor Reid is an associate professor of law here at Colorado Law and the director of Telecom and Platforms Initiatives at Silicon Flatirons.

[00:01:31.71] He previously directed Colorado Law, Samuelson-Glushko Technology Law and Policy Clinic and was a staff attorney and clinical teaching fellow in First Amendment and media law at the Institute for Public Representation at Georgetown Law.

[00:01:44.04] His work on topics such as content moderation, platform coverage requirements, copyright and disability access in the digital age have been published by journals including The Stanford Law Review, The George Washington Law Review, and The First Amendment Law Review.

[00:01:57.24] This panel will feature Eric Alston, scholar and residence at the finance division of CU Boulder, Dr. Newton Campbell, director of space programs at the Australian Remote Operations for Space and Earth, Professor April Dawson, associate Dean of technology and innovation and a professor of law at North Carolina Central University, and Professor Helen Norton university distinguished professor and Ross Gerber chair in constitutional law at the University of Colorado Law School. I'm going to turn it over to Professor Blake Reid for the panel and to tell you more about our panelists.

[00:02:28.59] BLAKE REID: All right, thanks very much, Catherine.

[00:02:30.43] [APPLAUSE]

[00:02:33.26] And let's just keep it going for one second for our amazing students who do so much to help out with these events. This is normally where I extol the virtue of a student like Catherine and tell you all to go hire her because she's amazing, but I believe she is already employed. So instead, I encourage you to chat with her at the break and lament that you are too late.

[00:02:59.16] With that, I am delighted to turn to our panel today on the intersection of the First Amendment and AI speech. The last panel, we were talking about robots interpreting the Constitution. It's funny how the tables turn now. The robots are coming and asking for the Constitution's help. We will decide whether they might avail themselves of that.

[00:03:25.62] So the way we're going to do this, we're going to talk through four topics. The first one is going to be the nature of the AI speaker and speech. Then we're going to turn to the potential harms of AI-generated speech.

[00:03:39.86] We will then turn to the regulation of AI speech and what the First Amendment might have to say about that. And then finally, we'll turn to the Supreme Court's recent jurisprudence on the First Amendment and how that might cross-cut.

[00:03:56.10] The way we're going to structure this, I'm going to bounce back and forth between our two expert Con law scholars to start things off on each question, Helen and April. And then I will look down the table to our technologists and business experts for some retorts and rejoinders.

[00:04:15.90] I have confirm that no one on the panel is an artificial intelligence. You are getting 100% human speech here. So let's get into it. So, Helen, you have written some of the path-breaking work on the nature of AI speech under the First Amendment.

[00:04:36.13] But I was taught by the earlier sessions that actually we don't need constitutional interpretation to be done by humans anymore, we can just ask ChatGPT to do it. So I'm sorry to tell you, I

asked ChatGPT if it had first Amendment rights, if what it says is protected by the First Amendment.

[00:04:57.21] And it says, as an AI developed by OpenAI, I don't have legal rights or protections like those provided by the First Amendment. So I think we're done. But if you want to rejoin that, go ahead.

[00:05:12.85] HELEN NORTON: Thanks, Blake. Well, I'm going to proceed as if I didn't hear that. But in thinking about whether or not the First Amendment has anything to say about AI, I think it's helpful to start by thinking about why do we have a First Amendment. What is the work that we think and hope that the First Amendment does? What are the values, the purposes that we hope that it's going to achieve?

[00:05:36.58] And it turns out that there's multiple answers to that question. One possible answer is that we should protect speech because speech is a valuable end in and of itself, that speech is intrinsically valuable because it's central to our individual expression, to our autonomy, to our dignity, to our self fulfillment.

[00:05:56.47] And so if you take this view of why we have a first amendment, our own speech should be protected because it's fundamental to how we develop our identities and express our ideas. It's part of our being fully human, such that it's just wrong for the government to interfere with our humanity by regulating our speech.

[00:06:16.25] And if that's your theory of the first amendment, if that's your answer to the question, why do we have a First Amendment, what is it supposed to be doing, then speech is protected, even if nobody other than the speaker finds it terribly interesting or important.

[00:06:28.77] Speech should be protected solely because of its value to the speaker's own personhood. And if you take that view of the first amendment, we probably should not understand the First Amendment to cover AI speech. And your findings are correct.

[00:06:42.14] But there are several other possible answers to this question, why do we have a First Amendment, what is it that we hope that the First Amendment would do, why do we think speech deserves protection?

[00:06:52.61] And these are more instrumental in nature. They are arguing maybe not that speech is a valuable end in and of itself, but that speech is a means to achieving valuable ends, other important objectives. And that speech can be valuable in achieving those objectives regardless of its source as human or AI.

[00:07:11.48] What do I mean by these sorts of ends or objectives that speech might help us achieve? Well, one is the dissemination of ideas and knowledge through the marketplace of ideas.

[00:07:21.95] This is the proposition that speech is valuable because it exposes us to ideas, and facts, and opinions that help us figure out

how we want to live or help us figure out how we want to think about important problems.

[00:07:34.07] Another instrumental argument for protecting speech is that we should protect it because it's a means to the important end of democracy. And this is the idea that a healthy democracy depends on the exchange of free thought, and expression, and dissent, where people can discuss and disagree about public policy, and candidates, and the government, and can talk about this in very critical, as well as very celebratory ways.

[00:07:58.76] Now, those are all-- those three theories I've discussed so far are all what you might call affirmative theories of the First Amendment. In other words, they're affirmative answers to the question of why the Constitution protects speech, because they say speech is affirmatively valuable, it's affirmatively good, so we should take care to protect it.

[00:08:18.44] But there's another very different answer to this question of why the Constitution protects speech. And this is what some call a negative approach to the First Amendment.

[00:08:27.63] Negative First Amendment Theory doesn't focus on why, or when, or whether speech is terribly valuable. Instead, here, the idea is that the government is too often dangerous in its capacity to abuse its regulatory power.

[00:08:41.82] The government is at times self-interested, or biased in other ways, or just clumsy in its regulation of speech. And for these reasons, we should take care to limit the government's power to regulate speech because we don't trust it to do so competently or disinterestedly.

[00:08:58.35] And this is called Negative First Amendment Theory because it's all about a negative value, distrust of the government, the idea that the First Amendment maybe is more about our fears of the government than about affirmatively celebrating speech.

[00:09:11.36] So these are multiple possible answers to the question of why the Constitution protects speech. And the Supreme Court is interested in them at various times, and commentators and lawyers are interested in them very at times. They're not mutually exclusive. And it's very common for a judge, or a scholar, or a lawyer to embrace more than one of them at the same time.

[00:09:30.60] But once we start to think hard about why the Constitution protects speech, we can see that most of these values, these purposes are at least as important to listeners as they are to speakers.

[00:09:41.13] So in other words, we value these views and the Constitution protects expression because it provides certain affirmative benefits to listeners and not just speakers. So expression under these

views, including AI expression, is valuable when it facilitates listener's democratic self-governance.

[00:10:02.04] AI speech and other expression is valuable when it enlightens listeners by exposing them to a marketplace of ideas, and facts, and opinions. And AI speech and other expression is valuable when it informs listeners choices, and thus the exercise of their autonomy.

[00:10:18.94] In other words, AI speech can provide value to listeners just like human speech. So those are the reasons why the First Amendment can be plausibly understood to apply to AI speech.

[00:10:31.24] BLAKE REID: So, , before I open it up to the rest of the panel, just a quick follow up on your last point. How do we know, how do we evaluate, what's the baseline for understanding when AI has served listener's interest in the way that you just described? Is that something that we can measure? Where do we source the content of that inquiry? How do we know? How do we apply that?

[00:10:59.11] HELEN NORTON: It seems to me that in answering that question, that the source of the speech doesn't matter, and you could just as easily have said, how do we know that any speech-- how do I know that your speech is terribly valuable?

[00:11:11.28] BLAKE REID: It's pretty straightforward. But go on.

[00:11:15.10] HELEN NORTON: I don't know that it lends itself easily to metrics. It's ultimately up to the listener to judge the value of the speech. And facts, and opinions, and ideas can be useful to them in terms of affirmatively shaping their behavior and their ideas, and they can be helpful to them in finding something that they want to push back on.

[00:11:33.47] And this is part of the idea of the marketplace of ideas, is that the speech that repels us may be as valuable as the speech that attracts us because it helps us identify what's good and true or wise, at least for us.

[00:11:45.72] BLAKE REID: With that, happy to open it up to the rest of the panel. April, can I come to you for a quick reaction? Then we'll bounce down the way.

[00:11:53.93] APRIL DAWSON: Yeah, so first, thank you for having me here. And , thank you for that wonderful description of the affirmative rights, the negative rights. You framed that so wonderfully.

[00:12:08.29] And as you were talking and I kind of knew where your conclusion was going, I wanted to get your thoughts on-- typically when we're thinking about the First Amendment, it's an individual person who has that right.

[00:12:22.73] So if AI ChatGPT, there's a First Amendment right for that information to be put out, in the event there's regulation of that, who would assert the First Amendment right? How would we see that?

[00:12:38.14] HELEN NORTON: So lots of good questions about operationalizing the idea. And that's maybe a set of pragmatic reasons what if if you're on the fence, you know, you can construct arguments for why AI speech might get First Amendment coverage in you're constructing arguments for why a speech might not want to-- those pragmatic limitations might be a reason to go on the side of, well, maybe it's more work than it's worth. But it's probably is worth the work.

[00:13:04.40] And it's definitely true, Professor Dawson, that you're anticipating that we would probably have to be more creative about our doctrine in order to operationalize that. But it seems like those are doctrinal problems that could be solved.

[00:13:18.68] We could, for example, trace the duties and responsibilities that come with speech back to their human owners, whoever unleashed them. I think that's the most common way to think about the possibility of operationalizing it.

[00:13:35.80] There's some folks-- I don't know that I'm quite ready to go there, but some folks are saying, well, as AI gets more and more advanced, AI itself might be responsible for maybe being enjoined if it has been found to violate folks-- if it's found to defame folks, or to threaten folks, or to steal from them.

[00:13:59.96] And this is hard for me to imagine, both how they could respond to an injunction and also maybe how they could pay a judgment. So why the more attractive, I think, idea to most folks right now is to trace it back ultimately to a human owner, or operator, or prompter.

[00:14:15.97] BLAKE REID: So I know we're going to go down the rabbit hole on regulations that do that tracing of liability back to the legal owners. But before we move on from this notion of treating AIs as a distinct entity onto themselves, I want to go down the way. And I'm going to start with Eric, and then come to Newton.

[00:14:34.72] Eric, I know you had some notes on our discussion before the panel about the cognition of AIs and what they are doing. Is there anything to be gleaned from thinking about AIs as speakers separate from the notion of what they output as speech?

[00:14:49.79] ERIC ALSTON: Yeah, absolutely. And I think a common understandable misconception from the strength of First Amendment protections that speech does enjoy is that, in some sense, all speech has value.

[00:15:04.81] And I'm not sure that's the case. I've encountered, and frankly, I'm known at the engineering firm I consult with as GPTE, in the sense of, I'm known to go on far too long.

[00:15:15.66] But I've been thinking a lot about, and I'm working on an article that might actually appear in this special issue, about the limitations of AI for the process of adjudication, what it is essentially to judge something.

[00:15:30.92] And I actually think most of the benefits that were described about why we protect speech in the first instance as interestingly linked to my problems I have with AI as any form of judge.

[00:15:42.85] I think AI surfaces existing information remarkably effectively, setting aside the issue of hallucinations, which are pretty well understood as a problem. But I think the notion that hard cases make good law, and that's a very debatable statement itself, but essentially, the higher courts have to adjudicate really hard cases.

[00:16:07.45] And ultimately, they have to weigh trade offs in ways that I call commensurate the incommensurable. Now, I've written about rights, trade offs in Ecuador's constitution, for example, but I think the problem is much deeper than that.

[00:16:22.01] And I trace it all the way back to the parable of King Solomon, which is, I think a lot of people who think, wow, what a clever judge he was in getting the revelation of the right answer, are missing the deeper point, which is King Solomon was presented with a case where one party had to lose completely. It was a binary outcome. One party got the baby. You could not split the baby.

[00:16:48.47] And so for me, the essence of some of our most important cases are ones where a judge has to use something fundamentally human, sort of, and ultimately something that cannot be distilled too easily commensurable quantities that you just weigh off against one another. When the answer is obvious, often the parties settle just out outright.

[00:17:12.29] And so for me, my issue is, is a lot of these purported benefits, and benefits I believe in, to be clear, of speech, I think rely on an expression surrounding the harder issues that mankind faces, including in our own self-governance.

[00:17:27.86] But the marketplace of ideas also hinges on trading off between areas we're uncertain about. You don't see people calling the marketplace of ideas, arguments that gravity doesn't exist or that the Earth is flat. It's thought to be that the issues over which there is substantive disagreement among members of society are ones in which it is beneficial to speak upon.

[00:17:51.65] But my issue is, the way an AI algorithm generates speech is, it is not weighing these really hard, I think, impossible for a

machine to effectively adjudicate trade offs. It is surfacing existing information and perspectives in a replicable way.

[00:18:10.28] And so for me, I view some issues with the way in which a large language model is generating an answer. That feels very foreign to me relative to the King Solomon, even King Solomon in our own minds, in deciding how we want to proceed on a particular issue of personal identity.

[00:18:30.59] And so for me, I view something essentially different in the way large language models are generating information and the way in which our human cognition generates speech in the important areas in which it does receive protection under the First Amendment.

[00:18:47.11] BLAKE REID: , I know you'll want to have a rejoinder to that. But before we do that, I want to go to Newton for his reaction.

[00:18:54.08] NEWTON CAMPBELL: Sure. I don't know how often you've been to Silicon Valley, but I definitely have encountered plenty of humans that are surfacing what seems to be complex ideas by just regurgitating what they've heard.

[00:19:06.85] So , as I was thinking about this topic as well, pretty much all of the things that you stated really gelled with me when I think about why one would want to protect AI speech.

[00:19:22.97] And really, for me, that marketplace of ideas really hits home at the diversity of what you can see by having something that is pulling in that information, but synthesizing it, and it has the ability to synthesize it in a near-infinite probability distribution space, is valuable inherently to the marketplace of ideas.

[00:19:45.18] When I'm thinking about this topic, I don't apply some form of agency to AI, ironically, given the AIs are agents. I don't apply some sort of agency. All I see it as is an ability to help humans in different ways, whether that's positive or negative, whether that's human beings wanting to execute something positive, like looking at climate change, or creating vaccines, or sometimes in the negative case, subverting democracies, things like that.

[00:20:20.14] The idea is to present humans with additional knowledge, additional expression. And I only see regulations or restrictions on that as regulations or restrictions on human beings having that knowledge and expression.

[00:20:38.12] So I guess that's really all I have to say about that. I acknowledge the dangers of it. But at the end of the day, for now, it is still made to serve humans. You're not seeing an AI agent that's out there in the wild unowned by anyone. A human being had to let that loose at some point.

[00:20:59.92] We will get there at some point. But for right now, the end goal is to serve human beings and human purposes in some way.

And I think from a tractability standpoint-- the tractability of the problem, it best serves us to figure out, all right, how do we trace responsibility back to the right humans when things go wrong?

[00:21:21.31] BLAKE REID: Helen, you're rejoinder before we move on to regulations on the role of humanity in constructing speech as such under the First Amendment.

[00:21:31.60] HELEN NORTON: Just two quick. I don't know that they were rejoinders, but responses. I mean, in part, maybe we're not being accurate when we talk about this as I speech, I mean, maybe at least now, because you could trace it to the human prompter if may be.

[00:21:51.23] But what it's doing so far is generating something based on what humans have said, or thought, or written so far. So it's not so clear that it's easy to disentangle AI speech from human speech, which is why I'm agnostic as to whether or not we should protect what we think is AI speech under the First Amendment.

[00:22:11.81] I think you can construct plausible arguments for why you can, and I think you can construct plausible arguments for why you can't. What really matters to me, and we'll talk more about this when we get to regulation, is a doctrine that focuses on human listeners, because I agree that the whole point of AI, whoever you think is generating it, is to serve human listeners interests.

[00:22:33.68] And as you'll see later, I think we should understand the First Amendment to protect speech when it serves human listeners interests. And I think we should understand the First Amendment to permit the government to regulate AI speech when it interferes with human listeners interests.

[00:22:48.44] BLAKE REID: So to transition us into that conversation about regulation, April, I want to come to you next. We have heard a lot of discussion today, including on the first panel, about the kinds of harms that AI speech might cause, both to listeners, to the public, and we think of other ways we might conceive of harm.

[00:23:12.62] April, what are the kinds of harms that you see as being most salient with AI-generated speech? And how do they differ, if at all, from the corresponding harms of human speech?

[00:23:25.79] APRIL DAWSON: Yes. So of course, the most obvious one is hallucinations. So we have misinformation. We have seen this. I'm sure everyone has heard about the attorneys who have, without thoroughly reading the cases that ChatGPT provided, included them in their briefs.

[00:23:49.39] There have been far too many lawyers who have just taken the information that ChatGPT provided and included it in the brief. And so these are perfect examples of hallucination and how these hallucinations can cause harm if the user is not diligent.

[00:24:08.50] So when hallucinations rise to the level of causing specific harm or causing defamatory information to individuals, then we can see another level of harm. So not just misinformation, but defamatory information.

[00:24:27.64] And there's actually a case out of Georgia Walters v. OpenAI. And this was a case where you have an individual who was a DJ in Georgia, and a reporter who actually knew this particular individual did a search on a lawsuit that had been filed against this company and asked ChatGPT for information about this particular suit.

[00:24:53.27] And ChatGPT came back and said Walters was actually the defendant in this suit. He was being sued for embezzlement when he was an employee at this company.

[00:25:05.00] Well, Mr. Walters was never an employee at this company, did not engage in any embezzlement at all whatsoever. But this was information that ChatGPT provided to this reporter.

[00:25:16.56] So the reporter didn't do anything further with that information, didn't further publish the information, did communicate with Mr. Walters. And Mr. Walters has filed this defamation action against OpenAI.

[00:25:30.35] So this is a really interesting case because it allows us to think about, if we do have some agency, either to the technology or to the company that has created the technology, what does defamation look like in this space?

[00:25:47.18] So OpenAI, they filed-- they requested removal to federal court. They ultimately withdrew that. But in their pleadings to the federal court, they actually filed a motion to dismiss.

[00:25:59.78] The case is back into state court. They haven't yet filed a motion to dismiss within that suit. But in looking at the federal motion to dismiss, it's very insightful. We know what OpenAI's arguments will be.

[00:26:12.14] So the first argument is that the content that was generated by OpenAI is not defamatory as a matter of definition. ChatGPT, any of these generative AI tools, particularly if we're talking about the text tools, they are not providing facts. That's not what they do.

[00:26:36.44] They are coming up with, based on statistics and probabilities, what's the next best word based on the prompt that they were given. So it's not facts that was being generated, it was the pattern based on the prompt that was given. So OpenAI's argument is, this was not meant to be factual in the first place, therefore it could not be defamatory.

[00:27:02.69] The other arguments that are raised by OpenAI is that there was no publication. So of course, with defamation, there has to be some publication. This went to one specific individual.

[00:27:14.45] It's not as though you have a newspaper that is publishing something or a book that's being published and being given to a whole swath of people. This is one individual who was requesting information.

[00:27:27.93] It's also unclear, because we know how these tools work, if someone else put in the exact same prompt, if they would have gotten the exact same result. So OpenAI's argument is there was no publication, it went to one person. And that person communicated it to Mr. Walters, but there was no mass communication to it.

[00:27:47.60] The other argument was that because Mr. Walters is known within his community, he's a DJ, there needed to be a demonstration of actual malice, because he's a public figure.

[00:28:01.89] And again, this is really interesting because when we're thinking about the generation of this text and we're thinking about if there is intent on the part of the technology, there is not.

[00:28:13.11] These are technological tools that are being used to help individuals so we don't have that intent. So actual malice, you communicated this information knowing it was false, or with reckless disregard as to the truth or falsity.

[00:28:29.65] And in this situation, again, kind of understanding the underlying technology, ChatGPT doesn't even know who Mr. Walters is. It looked at the information that was out there based on the training data, it may have even pulled some information from the internet, but there was no understanding. It's just pulling together text based on probabilities. So that's OpenAI's argument there.

[00:28:54.66] The other thing that I wanted to touch on before we get some other reactions is when we're thinking about artificial intelligence, and not just generative AI, but any type of artificial intelligence that is allowing for the communication of information at scale and at speed, we have to think about harm that may be caused within elections.

[00:29:23.63] So 2024, there are over 600 countries that will be engaged in elections this year. And when we're thinking about, of course, deepfakes, when we're thinking about sexual images, when we're thinking about audios, all of this can give rise to an acceleration of harm that has always existed.

[00:29:48.50] So when we're talking about misinformation and disinformation in the election context, that is not new, that has always existed. But if you are able to, as was the case in New Hampshire in January, create an audio file that is exactly President Biden's voice because they were able to pull that information and use a tool that's

widely available to have him say to New Hampshire voters, you don't have to vote in the Democratic primary because it's not necessary.

[00:30:21.75] It's only going to help the Republicans. Don't vote. And I don't know how many folks have actually heard the audio. But there's no way that you can, just based on listening to that, know that it is a deepfake.

[00:30:36.05] And so when we think about AI and the harms of AI, the fact that there can be a significant impact on the misinformation, and disinformation, and manipulation in the electoral context, that raises a lot of concerns.

[00:30:53.40] And the last point that I'll make with respect to disinformation and manipulation is to focus on pornographic videos. The vast majority of victims, when it comes to that type of generative AI, it's women.

[00:31:10.65] So when we think about female candidates, when you think about election workers, when you think about the threats, and harms, and intimidations that come from that type of threat, those that are being targeted are women, people of color.

[00:31:26.15] It happens at the National level. We know we will hear about it. There will be actions that are taken. So we saw this with the Taylor Swift situation. Well, if I'm running for a city council or for school board within my small little town in North Carolina, it is so easy for people to target me as well, but CNN is not going to pick that up necessarily. So these are some additional harms when we're thinking about AI and harms that we need to think about beyond just First Amendment rights.

[00:32:00.62] BLAKE REID: Well, thanks, April. That's a terrific taxonomy. I mean, I think you teased out both sort of retail harms, and then maybe you call wholesale or distributional harms. And I think we can pick both of those up. But I want to zoom in really quickly on the retail piece.

[00:32:18.13] And Helen, I come to you with a question about defamation, which is, it sounds inherent in the arguments OpenAI is making in that case that April just described, there are a couple of things emerging.

[00:32:35.25] One, they're sort of saying, the outputs of this are definitionally non-speech. There are some-- this is a product of math. This is not ontologically what we would think of as speech.

[00:32:50.47] And then the second, I guess, thrust of it is to the extent there's some speech there, that's arguably coming from the user. This is a mirror on what the user is feeding in, and maybe on what we have trained the model with. This is not us.

[00:33:05.98] And obviously, this is a familiar refrain in technology law debates. The intermediary says, not our fault. What do you make of those arguments given your conception of AI speech?

[00:33:19.07] HELEN NORTON: Yeah, I so appreciate those examples. And as you were talking about hallucinations and deepfakes, I was thinking, well, those are both types of inaccuracies. They are types of false or misleading speech, which is certainly a type of harmful speech that humans are very capable of producing.

[00:33:37.48] But you're helping us understand that, at least under some circumstances, AI speech can cause harms that are distinct from the harms caused by human speech, at least maybe exacerbating the harms that we see with human speech. So the harms are at least different in degree, and maybe sometimes they're different in kind.

[00:33:56.44] And so I was interested as you walked through OpenAI's defenses. I was trying to figure out in what ways do those defenses map on to defenses that we see in traditional defamation cases involving human speakers and in what ways does the AI speech invite brand new defenses that are unique to AI.

[00:34:15.70] And my intuition is that, even though I'm open to the possibility that AI speech deserves some First Amendment protection, to the extent that it's valuable to listeners, I just can't imagine any circumstance under which we would give AI speech greater protection than we would give human-generated speech.

[00:34:32.05] And the law of defamation, I think, is a terrific example. The law of defamation is a balancing. The tort of defamation seeks to make the victim of defamation whole for the very serious harms to their reputation that they've suffered while balancing those protections by crafting rules that ideally will not overregulate speech, such that it chills speakers from engaging in valuable speech, like valuable criticism of the government, where they might have some inaccuracies.

[00:35:00.13] And here in this particular case, the human tort plaintiff, defamation plaintiff has the same sort of reputational harms that we see in traditional defamation cases, but the AI speaker does not have the threats to chilling that we worry about in traditional defamation law. So long story short, I am unsympathetic to OpenAI's defenses.

[00:35:20.59] NEWTON CAMPBELL: Sure. So I'm going to ask what's likely 1L question here. If I have an engineering textbook, OK, If I author an engineering textbook and there is an uncorrected equation in that textbook, and an aerospace engineer designs an aircraft wing based off of that uncorrected equation, whether that equation was used for the testing of that aircraft wing and deceived them into thinking, oh, this aircraft is ready to go, that plane gets into a crash. Is the author of that textbook liable? This is not a snarky question. I have no idea. I'm a technologist.

[00:36:00.00] BLAKE REID: Defer to our Con law experts on that one.

[00:36:02.20] HELEN NORTON: Oh, I was looking for the tort law. Yeah, I mean, potentially negligence. If you know that this textbook is being used for this purpose, and it's reasonably foreseeable that people will rely upon the equation, and so you might be able to establish liability there.

[00:36:19.66] But there's also another issue, which is causation. So should it be the person who did the textbook or the ultimate builder of the aircraft. Isn't it reasonable for them to double-check?

[00:36:31.98] NEWTON CAMPBELL: I would argue it's the latter, especially in an arena where everyone has Lagrange equations. There's a million textbooks you can get, this is just one. So with the same not then apply even in the OpenAI case?

[00:36:45.79] You could have gone there for knowledge, but you could have also gone to some of these other platforms for knowledge, you could have gone to other data sources. Is that not something that we're going to consider when we're going to a group like OpenAI and saying, hey, they're liable? Is that-- just again, from a legal standpoint.

[00:37:07.30] HELEN NORTON: From a defamation standpoint, I think that doesn't matter. So if The Washington Post defames me, I can still go after The Washington Post, even though The New York Times says the nicest things about me. It's the fact that The Washington Post, I would go after them as the defendant.

[00:37:22.19] NEWTON CAMPBELL: OK, OK. Yeah, I mean, that I think kind of gets to some of the core of who effectively takes the blame in some of these cases though.

[00:37:35.94] If we have an AI that's telling you that this information about a journalist, or like you said, stating something as fact, when it's not really fact, it's kind of up to the end user, I would assume, to say, hey, you know what? I'm going to actually go check my sources.

[00:37:53.96] And within journalism, there should be some sense of journalistic rigor, if you will, to actually address some of those issues. Again, that was just an area that, I think from a First Amendment standpoint, I just wanted clarification on because I'm not familiar with defamation law at all.

[00:38:13.15] With that said, there's a number of cases that I've executed that fall along the same lines, where we are having an AI operate and tell the end user something. But at the end of the day, at least for right now, the end user is responsive.

[00:38:30.77] So we've had cases back in my DARPA and DOD days, where we would have AI assess the power grid, look at the power grid. If there's a massive nationwide blackout, it would actually tell you-- or an attack, it would actually tell you, well, here's how you cut off parts of

the grid, here's how you stave off parts of the grid. A very dangerous operation.

[00:38:52.01] But at the end of the day, because we raised that question with the military, it's like, hey, we're telling you this. Here's how this can go wrong. Here you go. At the end of the day, it was up to them to go, all right, this is what this AI is going to actually tell us, but we need to do our due diligence every step along the way, just in case this thing is wrong.

[00:39:16.14] And I think that from the perspective of a group, particularly like OpenAI, if we are going to do any sort of regulation, I think it would likely need to come with that clause that, hey, this could be inaccurate, which I think most people inherently know and I think might be buried in a terms of service somewhere. But I mean, it kind of seems like that overall is the solution in cases like that.

[00:39:41.11] BLAKE REID: [INAUDIBLE] want to pick up this question of responsibility allocation as we move along here. But Eric, I want to invite you in for reactions first.

[00:39:49.12] ERIC ALSTON: Yeah, and I may well provide a bit of a prelude for responsibility in the sense that I absolutely agree in principle that any speech or content generated by AI should not receive greater protections than that we already enjoy.

[00:40:05.36] And in particular, the harms listed, I agree are real harms, non-trivial harms. But I do think that effective regulation requires a question of who should be held liable and what is the likely deterrence effect on those you're intending to hold liable with respect to a particular regulatory intervention.

[00:40:26.93] And so I actually think resolving some of the speech questions posed by AI may implicate two other areas that I think are very unsettled on a constitutional footing that the digital age has brought forth at an exponential level, that of encryption and that of open source software.

[00:40:46.16] Because the problem being is I see two natural alternatives if we're not holding the machine liable and we're actually holding a human individual or a corporate person liable for speech created by an AI agent, and that's the developers of the company that developed the AI or the individual user putting a prompt in.

[00:41:07.04] But the reason that I bring up encryption and open source are, those have deep implications for the effectiveness of either of those as an appropriate locus of regulatory liability.

[00:41:17.51] If I'm an individual user given access to easily available encryption tools, I can spoof my identity in so many different ways that truly tying the output of an AI algorithm to my individual identity will be very hard.

[00:41:34.40] You might answer, hey, you shouldn't be able to encrypt your digital identity to that level, but that's precisely why I'm saying this is nudging up against encryption if we want to hold individual prompt, those who input a prompt that results in sufficiently harmful outputs.

[00:41:49.85] The problem with holding companies liable is, I think the cat's already out of the bag in terms of these algorithms have been open sourced, including by Meta, and are likely to continue to be so.

[00:42:02.66] And so there's a very interesting question lurking, which is, is code speech? Whether or not it is the status of open source software has a lot of implications for speech questions. And several courts have held coded speech in limited instances.

[00:42:21.54] Don't take away from my remarks that code is unambiguously speech under the jurisprudential definition of the US Supreme Court, but some courts have held a few steps towards that direction.

[00:42:33.48] And to me, the problem is, is if you regulate the companies that are not open sourcing their models, then that just means that the open source models will be the source of most of the harmful AI-generated content.

[00:42:47.06] BLAKE REID: So, Helen, I want to come back to you, and I want to ask a slightly different question than we had planned on, because I think we've shifted to this question of allocation of responsibility. I think we can back into what the First Amendment has to say about it.

[00:43:01.88] But I guess I want to throw one other piece on the table in addition to the real-world complexity, the technical complexity of open source models being out on folks laptops, on encryption, on the difficulty of watermarking.

[00:43:17.25] I might throw in a legal difficulty here as well, which is Section 230 of the Communications Act, which at least if you believe in its most ardent proponents, might actually afford greater protections to an intermediary that is hosting an AI model than we might otherwise afford if we treat it as a mere sort of mirror on its users speech.

[00:43:45.96] So that might provide an additional problem on top of the First Amendment. So Helen, with all of that, engage with it what you will. What is the First Amendment have to say about this mess?

[00:44:00.16] HELEN NORTON: Well, my first point is relevant to what Eric just said, in that some, if not a lot of what AI does I think can plausibly be understood as conduct rather than as speech that's protected by the First Amendment.

[00:44:16.07] In other words, I think that we could choose to understand algorithms and code as doing things. They are a set of instructions to machines rather than saying things.

[00:44:27.32] And if you take that point of view, AI might be considered conduct rather than speech. And that changes the First Amendment analysis and it makes it much easier for the government to regulate free from First Amendment obstacles.

[00:44:39.25] But even if we were to conclude that what the AI outputs, what AI is producing is speech or includes speech that's covered by the First Amendment, that doesn't necessarily require us to understand the First Amendment to forbid the government's regulation of AI outputs to address harms.

[00:44:57.54] And as I signaled before, I think we should take a listener-centered approach that understands the First Amendment to protect AI speech when it furthers listeners interests, and to permit the government's regulation of AI speech that frustrates listeners interests.

[00:45:12.66] Now, sometimes speakers and listeners interests are fully aligned, they're on the same page. When speakers want to speak and listeners want to be spoken to, they want to hear those speakers. But that's not always the case. And sometimes we have to choose between the First Amendment interests of speakers and listeners when those interests collide.

[00:45:31.32] And this is the case, for example, when speakers want to talk to listeners who want to be left alone, or when speakers tell lies and listeners want the truth, or when speakers want to keep secrets and listeners want accurate disclosures.

[00:45:43.04] In those situations, we've got hard First Amendment problems because we have to choose who we're going to privilege, who's going to win, speakers or listeners, when push comes to shove.

[00:45:51.53] And First Amendment law sometimes permits the government to choose listeners, especially in situations where speakers and listeners are not operating from positions of equality. For example, where speakers have more information than their listeners, or more power than their listeners, or maybe both.

[00:46:07.20] And the classic example here is commercial speech, and this is speech about the terms and conditions of a commercial transaction, speech about the price, or the quality, or the attributes of the dangers of commercially available goods and services.

[00:46:20.27] And the Supreme Court has held that commercial speech is worthy of some First Amendment protection because it's often valuable to consumers, to listeners who are making choices about what to buy or whether to buy.

[00:46:30.81] But the degree of that protection, again, turns on whether the commercial speech furthers listeners interests or interferes with them. So under this doctrine, commercial speech that is false, or that is misleading, or is related to illegal activity, first, it's not valuable to listeners. It frustrates their informational interests.

[00:46:48.66] And so the court has held, the government can completely forbid false or misleading commercial speech without raising any First Amendment problem because it's not helping listeners. And for the same reason, the government has substantial leeway to require commercial speakers to disclose accurate information about their goods.

[00:47:04.95] And the most famous example is the Surgeon General's Warning on cigarette packaging and advertisements. Congress has required the manufacturers and advertisers of tobacco products to disclose to their listeners, to consumers this particular health hazard.

[00:47:18.21] And once you know that, you can see it everywhere. It's on your cereal boxes, right? The federal government requires food producers to disclose nutritional information because it's a value to consumers as listeners.

[00:47:28.76] But on the other hand, commercial speech that's not false or not misleading is usually valuable to listeners. And so it gets a fair amount of First Amendment protection. And I think AI speech is in the same position. It can offer great value to human listeners, but of course, and Professor Dawson started us down this road, there's also a complicated and dark side to AI speech.

[00:47:49.56] And again, I think that we should understand the First Amendment to protect AI speech that's valuable to the listeners while permitting the government to regulate AI speech to protect listeners from certain harms. And I'll just give one example, and I bet my colleagues have a bunch of other examples. I'm going to flag manipulative AI speech.

[00:48:07.04] And by manipulative, I'm drawing on ethicist's understanding of manipulation that's unfair or wrongful behavior that involves a speaker's effort to shape their listeners decision-making to the speaker's advantage by targeting and exploiting those listeners vulnerabilities in ways that the listeners are not consciously aware of, so it's hidden, and in ways that those listeners could not easily become aware of if they were to try.

[00:48:34.10] And today's technologies, including AI, create opportunities for manipulation never before possible, both in degree and in kind. AI can identify users vulnerabilities by collecting and analyzing huge amounts of data about us, and then can develop interfaces and interactions that very effectively exploit those vulnerabilities to get their users to spend more time online, to shed more data, to buy more stuff.

[00:49:00.48] And I think that this sort of manipulative AI speech should be considered a category of unprotected commercial speech, just like false or misleading commercial speech, because all of this speech, manipulative, commercial speech, false or misleading

commercial speech undermines listener's interests rather than furthering it.

[00:49:16.70] So my aim here is to make sure that First Amendment doctrine retains a human focus, even if it is understood to cover AI speech, to retain a human focus by attending to the effect of AI speech, its value and its danger for human listeners.

[00:49:32.16] BLAKE REID: So lots we could pick up on here. But, April, I want to come to you. And please feel free to grab any of those threads that you'd like. But one that I'd be especially interested in, Helen has laid out a sort of a balance between speakers and listeners.

[00:49:48.40] It struck me in some of the harms that you listed earlier. There's yet someone else in the calculus from a policy perspective, which might be the subject of the speech. So when we're talking about deepfake pornography, for example, there might be a victim of that who is neither the speaker nor the listener, but might be the target of our regulation-- or might be the beneficiary, excuse me, of our regulatory intervention. What are your thoughts about how that ought to play into the First Amendment calculus when we've got someone else implicated?

[00:50:23.81] APRIL DAWSON: Yeah, I think, the thing about AI is it allows us to do that which humans are already prone to do in a faster way and in a broader way. And so I don't think that AI is the tool that is creating or helping to create at scale these pornographic images, changes really the First Amendment framework with respect to that.

[00:50:55.19] So if the First Amendment would not protect someone from either creating manipulative content or harmful content, it wouldn't even if AI was used to help them create that.

[00:51:10.34] I did want to pick up on something that Helen was saying about the manipulative nature of AI content if the user is using it for that purpose. If we go back to the elections and we think about Facebook and Cambridge Analytica.

[00:51:27.21] That was what the whole point was in order to gather-- the purpose of gathering that data, building those profiles, targeting specific communities for the purpose of manipulating what their thoughts were about the election.

[00:51:42.40] So when we think about different viewpoints, like who has these viewpoints, are there grassroots organizations that focus on this particular issue? There may not be, but you can manipulate people into thinking that it does based on the content that you produce with AI in these profiles that you build.

[00:52:05.11] So I don't think that when we're thinking about the harmful way that AI is being used, that it necessarily impacts how the First Amendment would apply in those specific situations.

[00:52:20.52] BLAKE REID: Let me go to Newton and then to Eric.

[00:52:22.99] NEWTON CAMPBELL: Sure. Just kind of picking up on that front. I think that is where a lot of the nuance comes in because I've had to work for a lot of advocacy organizations, where we do do a lot of the scraping, we do do a lot of the gathering and forming profiles of people, trying to lump them into clusters, but for things like how do you advertise to these folks for climate change, like campaigning and things like that.

[00:52:53.58] And so I guess I would say I agree with both points that I've heard thus far, but in terms of, I guess, the definition of manipulation, which I'm assuming has a long-standing legal definition, OK, maybe.

[00:53:14.83] I think that's where we would start to need to parse that, because I do agree that obviously that behavior is going to need to stop because the democracy is on the line. But where exactly we cut that off versus where a lot of the advocates are going, where a lot of the folks who are using AI in what I would call a positive way are going, we don't want to end up hindering them as well.

[00:53:39.91] ERIC ALSTON: Just to tack on. I think manipulation is a serious concern, but I think it should be weighed against the human ability to learn. And so I think the natural consequence of deepfakes, for better and for worse, will be greater skepticism as to video content that we encounter.

[00:53:59.34] And so AI, I think, is necessarily a step change in the technology of information production. But the storied anecdote of war of the worlds coming on the radio and convincing people that there was actually an alien invasion, this can be traced all the way back to the development of the printing press and the Catholic Church's concern over their tight control of existing information published in at least in the Western world.

[00:54:26.61] And so, again, that's not me taking a position of manipulation isn't a problem and we shouldn't do anything about it, but there is a countervailing tendency, which is as information becomes far cheaper and more effectively produced, it's used for both great and terrible ends at the same time.

[00:54:44.47] But ultimately, if we're harmed by those sets of ends, which are terrible, eventually we become skeptical of consuming certain sources of information. And so I think we may be deeply limited absent development of effective watermarks or other techniques.

[00:55:00.85] I think what the status quo equilibrium might be is, if you truly want to know what a politician is saying, go to a town hall or go to a debate, because otherwise, there's a much higher risk that the content you're consuming in a relatively unregulated online context is just AI-generated deepfakes.

[00:55:19.39] But if everyone knows that, that's an interesting equilibrium in which to assess what are the net social harms resulting from manipulation that we're seeing in the short-term, in great part, due to the extent to which the average listener is conditioned on an equilibrium that is no longer the case surrounding information production.

[00:55:39.81] BLAKE REID: I mean, I want to put that back to the panel, I guess, as a jump ball. That sounds like a terribly dystopian world, where we are flooded with information about the core function that we exercise as citizens participating in a democracy, and all of the information that informs our decisions on that front is fundamentally untrustworthy and we need to be really cynical about it.

[00:56:02.55] NEWTON CAMPBELL: We're already there.

[00:56:05.69] BLAKE REID: So we're in a very terrible place. Are we OK with that? I don't know.

[00:56:12.21] HELEN NORTON: Well, actually, I thought Eric was offering an optimistic point of view [INAUDIBLE].

[00:56:17.88] ERIC ALSTON: Yeah.

[00:56:18.27] HELEN NORTON: You had faith this morning, Scott Skinner-Thompson says he has faith and humanity to figure it out. I don't know that I have that faith, but it makes me feel better to know that some people do. So thank you for that.

[00:56:29.26] But there might be room for regulation if we're not comfortable with the status quo. And I think there's all sorts of reasons we should not be comfortable with the status quo.

[00:56:37.66] I focused my remarks on manipulative commercial speech, number one, because there's so much of it. And also, number two, because the First Amendment rules are much more listener-friendly in the commercial context than they are in the political context.

[00:56:51.73] But we can also find models for listener-centered frameworks in the political context. Citizens United. Whatever you think about it, every justice on the Supreme Court said, we are-- corporate campaign speech is valuable to listeners.

[00:57:08.98] We're not making an argument that corporations have autonomy or dignity interests themselves that are worth protecting. We're saying that corporations are major players in society, major players in the economy, and so what they have to say can be valuable to listeners. And where the majority and the dissent disagreed is whether or not unlimited corporate speech is ultimately valuable to listeners or actually detrimental to listeners.

[00:57:32.92] The valuable theory is the more speech is always better. And the detrimental theory is, well, actually, at some point, highly

resourced speakers can flood out competing views so the listener is exposed to an impoverished marketplace of ideas. And that's not the only example of a listener-centered framework in the political context, so I think there's also ways to be more interventionist.

[00:57:54.70] BLAKE REID: All right, I'm sufficiently heartened. Let's spend our last few minutes before we open it up to questions with-- I don't know if this is a lightning round, but let's call it current events. Why not?

[00:58:06.91] The Supreme Court is this term, and in recent terms has a lot of cases that might bear on this topic on its docket. April, I turn to you to walk us through what's happening at the Supreme Court.

[00:58:21.08] APRIL DAWSON: All right, so one of the cases that was argued-- actually there were two cases that were argued on March 18. And so one was Murthy v. Missouri.

[00:58:32.70] And so that's the one of two jawboning cases. And Eric, you mentioned Solomon. I think that whole jawboning terminology comes from King Solomon and something that he said at some point in time.

[00:58:45.90] And these cases really center around, can the government vigorously persuade, some may say coerce, individuals or companies in the case of Murthy, social media platforms to engage in some censorship or engage in some conduct that is based on someone else's speech?

[00:59:16.99] So in the Murthy case, this was when we had the Biden administration reaching out to Facebook, Twitter at the time, now X, other social media platforms, asking them, telling them, kind of threatening them, short of coercion if you ask the Biden administration, to take down posts that focused on COVID misinformation, at least according to the administration.

[00:59:48.45] And so the First Amendment is implicated because, of course, when we're talking about content moderation, we're talking about either deamplifying posts that are on these sites or actually removing them.

[01:00:00.85] And so Missouri and Louisiana brought an action against the Biden administration saying a couple of things. One, that the states actually have an interest in making sure that the administration doesn't undermine the First Amendment rights of its citizens.

[01:00:17.50] If you listen to the Supreme Court argument, I'm not sure that the justices-- any of the justices really bought that. And a lot of the discussion really focused on standing, like who has standing.

[01:00:29.02] The standing really centered around traceability. So one of the things that we know about these platforms is that they engage in

content moderation all the time. That's one of the reasons why they have Section 230 protection.

[01:00:43.14] They are going to be moderating, sometimes they're going to not take down something that might cause harm. Well, they don't have to worry about liability because of 230.

[01:00:54.04] So they're engaging in this moderation. And so the question with respect to traceability is, if someone's post was taken down, was it taken down because of the pressure from the Biden administration or because the platform decided on its own, this post needs to come down? So there's this traceability issue.

[01:01:12.52] The other standing issue that the courts spent a lot of time talking about was the redressability issue. So even if the court rules in favor of the plaintiffs and the respondents in the Murthy case, how do you redress it?

[01:01:28.19] The posts are already down. The post were very specific to a particular moment in time. So this pandemic that was going on, an election, election deniers. So a lot of the politics and a lot of the statements that were being asked to be moderated were specific to that time frame. They no longer exist. So there are these issues with respect to redressability as well.

[01:01:55.39] The other really important point that the justices spoke about during the oral argument was, at what point does the government's involvement, maybe even partnership with the social media platforms, make the content moderation a state action?

[01:02:15.28] So if we're talking about the First Amendment, the First Amendment only applies if we're talking about the government engaging in some conduct. If it is focusing on private conduct, then you have to show that private conduct has been converted into a state action. And so the question is, was that entanglement, if you will, did it rise to the level of actually being a state action?

[01:02:40.87] The other jawboning case, which didn't really involve technology, this case, it feels as though the court is going to rule in favor of the NRA, where you had the director of the New York Department of Financial Services advising, kind of threatening, persuading insurance companies and banks not to do business with the NRA because of opposition to the NRA's promotion of gun rights.

[01:03:07.54] That's a much easier case as far as the threats that were being made. It's clear to show the traceability, the redressability. So it will be really interesting to see what the court does with that case.

[01:03:22.25] The last case that I want to just highlight. So this is the NetChoice case. And so this is an interesting case, because with the Murthy case, you've got the federal government trying to get these platforms to take down specific content. In the NetChoice case, you have the state governments trying to prohibit these social media

platforms from taking down or engaging in certain types of content moderation.

[01:03:51.52] Again, the facts that gave rise to the NetChoice case was this perception that social media platforms have some animosity towards conservative leaning statements or conservative leaning expression, that viewpoint.

[01:04:10.17] And so two interesting cases, one out of Texas, you have the Fifth Circuit, which upheld the Texas law. And then you have the Florida case, where you have the 11th Circuit striking down the Florida law. So we've got this split in the circuits.

[01:04:27.10] And really the question comes down to whether the states are able to moderate content or restrict the moderation of content for the protection of the citizens. So again, we see the states positioning themselves, where they are protecting the First Amendment rights of others.

[01:04:50.19] Also, Helen was talking about the listeners, kind of the rights of listeners. We see this in the NetChoice case as well. So if the social media platforms are removing tweets and removing Facebook posts, there are those that might be interested in hearing that or seeing that. And so they're harmed by it as well.

[01:05:15.29] But again, it goes back to, are we talking about a private platform? Is this platform a common carrier such that the government can regulate ? It'll be really interesting to see what the court does with that case as well. So those are the three cases that the Supreme Court is wrestling with.

[01:05:33.36] BLAKE REID: Oh, and maybe just to put two recent examples, where this has blown up for artificial intelligence on the table. We have both the recent example with Google's Gemini Model creating diverse soldiers in Nazi regalia. And then we've got the Gab racist chat bot that is intentionally being configured to spew out a lot of racist ideas.

[01:06:07.26] And then calls in from both political directions have not manifested in legislation yet, but that might call on platforms to do different pre-prompting, different training, different moderation, both inputs and outputs to deal with that.

[01:06:23.29] So we might see these issues bubble up. With that, I want to come down the panel from the end, and start with Eric, then to Newton, and then to Helen for a final word before we go to questions.

[01:06:33.63] ERIC ALSTON: Awesome. And so I just want to double-click on, given the current events prompt, two areas where I think we are going to see definition that both implicates the future of AI models, but also a variety of interesting speech questions.

[01:06:50.02] The first surrounds encryption. And so I would encourage the interested to consider the Tornado Cash case currently. And setting aside any of your beliefs in terms of whether we have a right to economic and financial privacy in the digital realm, that case itself both implicates encryption rights, as well as open source questions, because the developer of Tornado Cash is being held liable for end uses of a piece of open source software-- well, the developers, I should say, are being held liable for the end uses of a piece of open source software that has implications for the effectiveness of regulatory interventions with respect to AI.

[01:07:32.66] But it also, to me, encryption in the digital age is an important shield that provides associational and speech privacy. And so there is a decent bulk of jurisprudence protecting associational and speech privacy in certain contexts, mind you.

[01:07:51.20] And so it's again, not absolute, but very interesting questions in the digital age for speech and AI raised by both encryption as well as the extent to which developers can be held liable for end uses of software that they let out into the wild, so to speak. I'll end there.

[01:08:10.79] NEWTON CAMPBELL: So all very fascinating cases. I think, when it comes to open source capabilities and even open use capabilities, like things that the things that OpenAI pushes out, I think it's important for us to remember why we started doing open sourcing in the first place.

[01:08:33.14] Open sourcing goes way back, but it got very popular in the mid to late aughts. And the overall idea, the overall goal wasn't for this thing to be completely ready for prime time, be completely ready for operation. The onus, in fact, if you were going to do that, was really on you and your organization.

[01:08:54.15] Really the idea behind it was I'm an organization. I'm a company. I'm a grad researcher who has a model that I think shows some promise or I have some code that I think shows some promise. And I want the community to chime in, tell me what's wrong with it, tell me what's right with it.

[01:09:13.44] Now, obviously the OpenAI example and even what Meta AI is doing, those are substantially different in that these companies are actually-- they're making money off of the fact that these models exist.

[01:09:30.39] They know that some folks are going to be depending on it, but they still do have those roots of, tell us how this is doing? That's what the little thumbs up, thumbs down button for ChatGPT is.

[01:09:44.61] And I guess we really do have to, at least in the current times, think of what the balance is between restricting certain things that these models can say or do versus leaving room for them to

actually get better. Because the purpose of that open sourcing process is for them to get better. They cannot get better without the public using them.

[01:10:08.44] And we do need to foster a public that, I mean, effectively is educated enough to be able to use it and use it responsibly. Should there be guardrails there, where we get into who's responsible and who's not? I don't want to dive into that.

[01:10:25.02] But I do think that it's important to remember why open sourcing exists in the first place. It was never to have-- to say that you have a ready-to-go product. It's to say, hey, I want you to test this out. I want you to try this out.

[01:10:41.95] BLAKE REID: Helen, a final word before we go to Q&A.

[01:10:44.82] HELEN NORTON: So as Professor Dawson was going through the list, I was struck by how a number of the cases involve First Amendment challenges to what we might understand as non-traditional forms of speech regulation.

[01:10:56.59] So the jawboning cases are complaining that the government's speech, the government's speech is usually exempt from free speech clause review. The government can usually talk as much as it wants.

[01:11:06.06] And your remedy, if you don't like what the government says, is to elect yourself a different, better government. But the government's speech by itself doesn't violate the First Amendment.

[01:11:14.83] And here, the government, through its speech, was trying to shape its listeners choices. And the question in these cases, the First Amendment question is, when, if ever, does the government speech step over the line from permissible persuasion to impermissible coercion? And the less the government can speak, that takes away a very commonplace, but non-traditional regulatory tool.

[01:11:40.57] And then the NetChoice cases have to deal-- it's a First Amendment challenge to state laws that try to regulate platform's self-governance. They're trying to shape the speech on their platforms through markets and norms. Through their terms and services, they're trying to shape their users norms, this is what we allow, this is what we don't allow, and here's how we enforce it.

[01:12:01.78] And through markets, if you don't like the rules, well, you can choose another platform. And we're going to learn from these cases whether or not that sort of corporate self-governance is going to be permitted as a non-traditional regulatory tool.

[01:12:14.55] And Blake, Eric's mentioned of open source reminds me of another tool that you might think of as a non-traditional regulatory tool in the speech realm, which is copyright, which is a funny little beast of its own.

[01:12:26.71] BLAKE REID: Yeah. And maybe worth a very brief mention here, just to say, there's a literature that copyright scholars have called the weaponization of copyright to use it as a regulatory model for harms, whether we're talking about privacy, or specifically in the context of AI to achieve some broader set of regulations that are beyond what we traditionally think of copyright as being afforded to structure transactions around the dissemination of creative works.

[01:12:58.12] And one of the interesting things about the ongoing litigation against many of the AI platforms involving copyright is, we actually are very unlikely to see First Amendment challenges. And in fact, some of this litigation is starting to advance.

[01:13:17.43] And that is because, under copyright law, we typically cash out First Amendment questions a different way. We question whether the use being made is fair, which prompts an entirely different set of questions and equities.

[01:13:33.34] And so I think there is a lot of interest and potentially hope from some quarters that copyright might be used effectively to structure some of what the platforms are doing or to regulate what some of the platforms are doing without having to worry about some of the First Amendment harms that we have discussed.

[01:13:54.13] With that, I am happy to open it up for questions. As our tradition goes, I'm going to come to a student first for questions. And there's a corollary to this rule, which is when it's late Friday afternoon and I see a number of great students who have hung out all day, I will call on a student for the first question, but I'll keep going with students as long as there are student questions to come.

[01:14:17.39] So I see one down in the front row. Catherine, as the microphone comes to you, I can only say, I don't think I have ever had a student in my office hours as often or as angry about the NetChoice cases as Catherine. I anticipate a question on that. Go ahead.

[01:14:32.98] AUDIENCE: Sorry to disappoint. I had a question for Professor Norton and Dawson. As we're thinking about regulating AI speech, I keep thinking about the Supreme Court's continued and increased adherence to the idea of originalism, and history, and tradition, and custom.

[01:14:52.67] And I'm thinking about the Steven's Holding, that to create a new unprotected category of speech, there needs to be a long-standing history and tradition. And to quote the NetChoice cases, not in their wildest dreams could the founders have imagined OpenAI. And so I'm wondering, how do you see the Stevens Holding complicating or playing into the regulation of AI-generated speech?

[01:15:21.24] HELEN NORTON: Yeah, that's a great question, Catherine. And my first reaction is it shows the limits of originalism as a tool for responsible constitutional problem-solving.

[01:15:33.21] But I would still-- I think there's ways to limit and distinguish. So Stevens-- United States versus-- there's a handful of categories of speech that the Supreme Court over the centuries has held is entirely unprotected by the First Amendment, defamation threats, obscenity, child pornography, a couple of others.

[01:15:49.54] And for the longest time, the court explained that these are unprotected because they're low value First Amendment value and high harm. They're not doing any First Amendment work, so the government should be able to regulate them.

[01:16:02.02] And then in the Stevens case in 2010, the court said, oh, no, no, no, no. We're not doing cost benefit analysis. We never did it. Even though they did. We're not doing cost benefit analysis.

[01:16:10.95] The only way you can identify a category of speech that's entirely unprotected by the First Amendment is to show that there's this long history and tradition of regulating it without triggering any First Amendment heartburn.

[01:16:24.20] So it suggests, and lower courts have looked at this to decide, well, I guess that non-consensual pornography is not unprotected speech because, of course, there's not a long-standing history tradition of regulating something that hasn't existed. But on the other hand, a number of those state courts to consider regulations of nonconsensual pornography have held that they satisfy strict scrutiny.

[01:16:47.15] So I think what Stevens will do is to require-- is to force courts to do workarounds around its holding and either characterize the government's regulatory target as content neutral and not content based, so lower levels of suspicion, or in rare circumstances hold it, and I do think it will be rare, holding that the government's regulation nevertheless passes strict scrutiny.

[01:17:09.67] APRIL DAWSON: Yeah, great question. And I think what I would add is, your question illustrates the problem that we have in the judiciary, the executive, and the legislative branches, which is technology is moving so quickly and our government moves slowly.

[01:17:28.88] I mean, even if we think about lawyers and law professors, this is a wonderful conference, but most law professors aren't thinking about these technological issues as much as we need to, especially since you and your colleagues, you will be dealing with this as lawyers. So we're very slow-moving as institutions.

[01:17:51.55] I think what-- having listened to the arguments, my sense is that the court will decide these cases, these complex technological cases on the narrowest grounds possible.

[01:18:04.94] So if you think about Murthy, I think they're going to rule on it on standing. That's an easy way to go ahead and resolve it and to give these issues time to marinate, to help the justices wrap their heads around these emerging technologies.

[01:18:21.84] They're doing better, but they don't live in this space either. So I don't think they will craft new frameworks or new rules. I think they will go incrementally. I don't know if that's to our benefit.

[01:18:39.40] And then I just want to very quickly tie that into Congress. Congress has had a lot of discussions and hearings. There was a Senate hearing just recently on AI in the elections. Great, great hearing. A lot of great discussion.

[01:18:52.61] But we can't just have discussions. I don't know when we're going to get legislation. And your question illustrates that point, that these things are moving very quickly, but we're moving very slowly.

[01:19:03.97] BLAKE REID: And just to tag on to that, and say that this is the copyright point, which is copyright law is the only real law on the internet. It's the only place we've applied regulation to both internet platforms and internet content for the better part of the last three decades.

[01:19:22.10] With that, any more student questions. And if not, I will go to the rest of the audience. I see a student of life, Mike [INAUDIBLE] right here on the right side. Over to Mike. And if you could wait till the microphone comes to you.

[01:19:41.35] AUDIENCE: Let's hear. One thing that often strikes me about AI in a legal context is that the presence of the word AI seems to act as a magic talisman to activate complexity that doesn't have to be there.

[01:19:54.04] So, for example, there's much made of the non-consensual pornographic images produced of Taylor Swift by AI. But if we went to Taylor Swift and said, fantastic news. We've looked into it. Actually, they weren't produced by AI at all. I doubt that she would say, ah, well, forget the whole thing then. I'm not bothered.

[01:20:12.83] And similarly, many defenses of AI are made by the fact that it's just wildly complex. But when I taught, I used to tell my students correctly that if we printed out all of the specs, and manuals, and so forth for the Boeing 747, it would weigh more than the plane.

[01:20:32.04] And yet, I bet that when Boeing gets sued about these problems they've had with their aircraft, they don't successfully mount the defense that, well, jeez, we got everything right except this one bolt, and this plane is just hopelessly complex. We can't possibly be held liable.

[01:20:49.41] If they build a plane that's too complex to operate, they're not allowed to have it. Problem solved. So why are things like the Swift case and the complexity defense for various things allowed for AI, but for nothing else?

[01:21:04.80] BLAKE REID: And Mike, that's a great question. I think it links to a discussion during the earlier panel about which takes primacy here, the technology or the law? And we might ask ourselves--

[01:21:16.95] AUDIENCE: I can withdraw my question.

[01:21:18.13] BLAKE REID: No, no, no. I'm [INAUDIBLE] I think it's important. Why does it matter that AI is so complicated? The law is allowed to have simple instrumental goals that can ignore that complexity. Reactions from the panel. Newton, you look itching to jump in.

[01:21:37.09] NEWTON CAMPBELL: I was just pondering. No, I mean, I see your point with respect to particularly the Boeing defense. And I think when certain things exhibit specific kinds of emergent behavior, you get past the point where it's just complex. And when they exhibit things beyond just normal use, now you start to get to the point where, OK, we need to actually do some regulation about this.

[01:22:08.33] Social media is not that complex of a tool until a million people are on it, then it becomes more of a complex tool, then it has an impact on the democracy, then we start to ask ourselves, should we regulate it? Now, I don't know if we regulate it via some mechanism through the First Amendment or we just regulate it through another mechanism.

[01:22:29.30] But I do think that it having such wide and critical use cases allows it to be a candidate for some kind of regulation. Now, what kind of regulation that means, whether or not that's regulation on the development of it or the use of it, I think we need to have some nuance there.

[01:22:56.65] But I mean, yeah, I guess I would agree with, it's not so complex that we shouldn't think about how to regulate it. But I do fall back to the notion that it is out there, particularly the ones that we're talking about, ChatGPT and things like that, not some of the other services and robust models, but some of those technologies are out there as open source capabilities that are effectively being tried out.

[01:23:27.31] And I think that if we're going to regulate anything, it should be about the messaging, the terms of that, the making people aware of its inaccuracies, and so on, and so forth.

[01:23:39.96] BLAKE REID: Let me grab the other half of Mike's question real quickly while I ask. When we evaluate the application of the First Amendment, does the Supreme Court do that in a technology neutral or agnostic way, or does it take into account the characteristics of the new medium? In other words, is the new attributes that AI has brought to the table, does that warrant some new consideration, or are we just old wine in new bottles?

[01:24:11.61] HELEN NORTON: Yeah well, the example of non-consensual pornography and the example of defamation, the

regulatory interest, I think, is often the same, regardless of whether the non-consensual pornography or the defamation is produced by a human or an AI source.

[01:24:28.76] So I agree the law shouldn't turn on the AI source of non-consensual pornography because the harm to the victim doesn't turn on the human or non-human source. There might be some situations where the nature of AI exacerbates the harm or makes a more challenging regulatory target, which [INAUDIBLE] models.

[01:24:52.70] And if that's the case, if we're seeing new regulatory models, that might require the court to be tech sensitive. But often I agree. Often these challenges do not require the court to be tech sensitive.

[01:25:05.00] BLAKE REID: With that, I'm getting the hook from our host here. So please join me in thanking our panel for the wonderful discussion.

[01:25:12.69] [APPLAUSE]

[01:25:14.21] Thank you, guys.

[01:25:18.05] HARRY SURDEN: Well, Thank you, Blake, and our panelists for a terrific panel to end the day. And I want to thank everybody more broadly for an amazing day. We had some really great insights about the intersection of artificial intelligence and law. I know I learned a lot. I just want to, again, thank a couple people who worked so hard to make this come together. Professor Suzette Malveaux, my--

[01:25:40.24] [APPLAUSE]

[01:25:41.36] --co-organizer here, the staff at the law school, Dean Ines, Brad Bernthal, the amazing team at the Silicon Flatirons Center, Shannon Sturgeon, Christine McCloskey, Nate Mariotti, and Sarah Schnittgrund,

[01:25:57.59] [APPLAUSE]

[01:26:01.36] Law school events coordinator, Lindley Bell.

[01:26:04.13] [APPLAUSE]

[01:26:06.28] The IT staff, who were amazing here, and the student volunteers, who were terrific, and the fellows. And last, but not least, our terrific panelists who came far and wide and shared their expertise. So thank you so much.

[01:26:22.45] [APPLAUSE]