# Audio file

[Veena, Sandra, Brent OII 16th Jan 25\_EDITED.wav](https://unioxfordnexus-my.sharepoint.com/personal/inet0394_ox_ac_uk/Documents/Transcribed%20Files/Veena,%20Sandra,%20Brent%20OII%2016th%20Jan%2025_EDITED.wav)

# Transcript

00:00:00 Veena McCoole

Hello and welcome to the Oxford Internet Institute podcast, part of the University of Oxford. In each episode we look at issues and developments in the digital world that matter to us all.

Today we're joined by professors Sandra Wachter and Brent Mittelstadt. Sandra is professor of technology and regulation at the OII where she researches the legal and ethical implications of AI, big data and robotics, as well as Internet and platform regulation.

00:00:27 Veena McCoole

Brent is professor of data ethics and policy at the OII, and the OI is Director of research. His research spans ethics, law and emerging information technologies.

I'm Veena McCoole, media and communications manager at the OII. Welcome to the podcast, Brent and Sandra.

00:00:44 Brent Mittelstadt

Thanks so much.

00:00:44 Sandra Wachter

For having us, thanks so much for the invitation.

00:00:47 Veena McCoole

So to start with you, Sandra, tell us about the current state of artificial intelligence regulation. Where are we right now as far as regulating AI?

00:00:54 Sandra Wachter

Yes, I think it has been a really exciting couple years, actually a lot of of different types of regulation.

Did find its way through the legislative process in the last couple of years. I think for our purposes the most important ones to mention coming from Europe at.

East is the Artificial Intelligence Act that is now being enacted, which will be a massive great framework that will create obligations for people who want to develop or deploy systems.

There will be a new liability directive that is supposed to give people recourse if they feel like that Nii has harmed them or damaged them in in a certain way.

And then we're also going to have two other frameworks. 1 is called, the Digital Services Act and one is called the Digital Markets Act. Both of those frameworks are trying to regulate big platforms and trying to make sure that what happens on the platform is ethical, is legally sound, protects human rights.

And also to make sure that the big players are not exploiting their market power. So making sure that newcomers, SMEs and startups also have the opportunity to enter the market. So that's what's happening on European level at the moment.

00:02:08 Brent Mittelstadt

And I can say something quickly about what's happening within the UK and the US.

So within the UK just a couple days ago, Keir Starmer announced that the UK was going to take a pro innovation approach to AI regulation. Which is interesting because the Conservative government before him said essentially the same thing it was meant to be very pro innovation.

And there was a sort of specific contrast drawn with Europe, where Europe is sometimes perceived as sort of anti innovation. The UK was going to take this pro innovation approach. So we'll see sort of what happens with that. We have AI bill introduced previously in the UK, but it didn't go anywhere.

00:02:43 Brent Mittelstadt

So maybe we'll see that again and in the US, so we have of course the executive order on artificial intelligence that was published under the Biden administration, legal scholars in the US are largely expecting the new Trump administration to repeal that or to cancel that. But we don't know what, if anything, it will be replaced with in the US.

00:03:05 Veena McCoole

And what are some of the most pressing challenges faced by AI regulators today? You touched on a few of.

00:03:11 Sandra Wachter

Them, yes, I think there are a lot of different risks and opportunities just to name, I mean three that I think are probably the most important ones that we have to deal with that are important for everybody.

Regardless of where our where AI is being deployed, one has to do with bias. Another one has to do with opacity and the third will be privacy implications and so the bias issues. Really one of the most.

Issues that we need to to think about when it comes to AI and it has just to do with technology itself, like machine learning and AI basically is looking at the past trying to predict the future. So it's being fit with historical data.

So historical data on who has been higher than the past, who has gotten a scholarship in the past, who had to go to prison, who was given a loan, who got healthcare, and when we think about who got healthcare, who was allowed to go to university and who was allowed to go, had to go to prison.

Very often, those decisions are not necessarily marked.

Good equality standards. Very often that's marked by inherent problems of sexism, ableism, heterosexism.

Racism. And so it's inherent in the technology that it will pick up those unfairness, those those unfair patterns, and transport them into the future. And so we're gonna have systems that make the same problematic decisions again.

Being sexist and racist when giving out loans or awarding scholarships or giving out healthcare because AI can only learn from what it has been given and so like bias, is always going to be a major issue. Similarly, that the problem of opacity. So that means that means that we don't fully understand how those systems actually work.

It's called the black box problem. That means that not even a developer really fully understands how and why AI makes decisions. So if you're awarding a scholarship, if you're giving somebody healthcare, if you're denying somebody a loan, why that was actually decided.

Is not really clear to people, and again, this is a really big, big issue that we need to to care about and the last one has to do with privacy because, you know, AI is kind of useless if you don't put data into.

And so as soon as you have data, you have a privacy issue and so there is a vast amount, vast amount of data that is being fed into this. And very often this data can give away very intimate details about your life that you're not fully aware of. So you might think you just open up an app or.

Firing up a web page on your on your browser and you're not aware that the keystrokes that you leave behind the search patterns how you interact with the Voice Assistant and the how, how you how you voice.

For example or.

Those kind of patterns give a lot of information about a way about you, so it might be that AI is able to infer your health status, your ethnicity, your gender, your sexual orientation, who you voted for, if you voted, how much money you have.

Just based on how you interact with digital technologies and I think people are often not aware that they giving up that data because AI is really, really good at finding out your deepest secrets by looking at data that looks very, very boring or trivial.

And so I think that's just really important to.

00:06:32 Brent Mittelstadt

In mind, yeah. So in terms of what regulators are facing sort of on the ground, I think.

One of the biggest challenges they face is political pressure right now, because there's so much political will sort of across the board, you know, within the EU, within the US, to be pro innovation.

And I think there's a push to, you know, for regulators to really allow companies to do things much more freely than maybe they've done them in the past.

To sort of accept innovation and try to regulate it after the fact rather than you know, trying to regulate it upfront or make sure that innovation is sort of aligning with our values as a society.

So there's that sort of resistance to political will that is needed, I think for effective regulation, but then also even if you have sort of all the will in the world, there's also the problem of translating, you know those high level regulations high level.

Values down into specific requirements. Things that a company can actually implement, you know, rights that a citizen could actually exercise.

A lot of that is being done through what are called technical standards. These are things that are being developed by these committees that are made-up of industry of civil society, but predominantly or industry led, where essentially they're setting their own rules.

There it's almost a form of self regulation. Technically it's not, but it feels very similar to that, just cause industry is such a powerful voice.

There and there remains a lot of vagueness and a lot of plurality within the standards that are being developed around AI. So it's almost like what regulators have done is sort of kick A can down the road to say, you know, what does bias mean specifically? What is a good explanation?

We're going to answer that through standards and then now that we're drafting the standards, it's like, oh, these are really hard questions, actually.

And we can't say universally what a good explanation looks like. What bias actually means, what is.

Unfair. And so now you have this vagueness that's built into them. That means if you're a regulator trying to enforce these things, you know, how do you actually hold a company accountable to a requirement that is inherently vague? It's a really, really difficult challenge, and they're trying to do this.

While also having the most sort of talented people available, being poached by industry who can pay much more and the sort of expertise you would need to actually, you know, audit these systems to actually understand what's happening within them, just not available at the sort of scale that the regulators would need to be.

00:08:53 Brent Mittelstadt

Really effective.

00:08:56 Veena McCoole

Right, so as you both were sharing those challenges faced by regulators, it occurs to me that many of these issues are not new, you know bias opacity. Don't we have laws protecting us from some of these risks already?

00:09:08 Sandra Wachter

Yes, I think it's a. It's an excellent question. It's also.

I think the guiding question that we have in our research group here because Brendan I and Chris Russell, who is also here at Oxford, we have had a research group since 2000 and and 18 and that is usually the guiding question that we ask ourselves aren't?

Something that we're already aware of, familiar with haven't regulator had to deal with that in the past already and that's usually the first step of of investigation that we do so in, in this case, you would think, oh, we have anti discrimination laws, nondiscrimination laws, because unfortunately bias and discrimination is not something that is new to society.

But the issue very often is with this and many others of the AI issues is that the law or the laws were written by humans for humans. So that means you think about how human acts, how a human is motivated, what their actions are, and you think about tools to prevent.

And punish them if they do something that is problematic.

Very often, how the harm occurs when AI makes decisions is fundamentally different than when humans do it, and so it's not easily translatable anymore. The law isn't easily constituted by more.

Non discrimination laws is one of the best laws ever written, and it's very, very clever.

But it is for good reasons, based on a complaint based system. So complaint based system means that I have to feel that something is wrong. You know, I'm being harassed. I'm being bullied. Somebody's fired on on fair grounds. I know that I get treatment that I shouldn't be getting and therefore I'm.

Enabled to bring a complaint complaint based system right.

So with the eye, for example, discrimination often happens behind your back without you being aware of it. So with the examples that I mentioned earlier, right, you just opening a web page, you're browsing for something to friends on Facebook that you have the things that you like, the group that you join, they already give information about you away without you being aware of it and information.

Being used to tailor their online.

Experience online world for you. Like everything that you see online is tailored to you, right? We don't have a shared reality anymore. You and I are not looking at the same magazine anymore. I see different search results than you do. I get different prices than you do.

And so it's really, really hard for me to see if you're getting better prices or worse prices or if you're getting better job offers or less good job offers. And so this comparative element that is really needed to bring a complaint is completely.

Eroded. And so a complaint based system doesn't really work anymore. If discrimination really happens behind your back. And similarly, it's going to be super, super hard to prove that case, right? With usual discrimination cases that are already hard to prove.

You you might be able to show that there's problematic behaviour, so for example somebody saying oh, you shouldn't work.

Headdress at the workplace, which is really clearly linked to freedom of religion, for example, right? But a person asking you during a job interview if you like to colour?

The Reds might sound like it's odd, but it doesn't make your alarm bells ring. But there could be an algorithm that is able to find a correlation. Not causation, but correlation between liking the colour Reds and being gay, or believing in a certain God and all of a sudden you have this proxy that humans would never think to use.

But indirectly, then act as a discriminator again without having enough proof to be successful in court or AI just comes up with completely crazy groups to make decisions based on, so we make decisions based on things.

That makes sense to us in in a world, but AI finds correlations and might be.

Grouping you in a way that the law doesn't fully recognise as a group that needs protection like we know that gender, ethnicity, ability, religious beliefs, age, those are groups that we want to protect. But what?

If I told you that if you're applying for a job and you're using, for example.

Internet Explorer or Safari to submit your online application. I can tell you that you're gonna be less likely to be successful, so pro tip if you're applying for a job, use a browser like Safari or Chrome because your chances of getting the job are much higher. Like that is something that has been done for quite some time, but obviously.

Internet Explorer's are not or Internet Explorer users are not protected groups under the law because we never had to be.

And so all of a sudden we have those new decision criteria groupings that the law doesn't know anything about because it never had to be protected in the past, right? And so we have discrimination, which is not a new problem, but how it's being created, how it's being motivated and how it.

Exemplifies itself. That's the thing that is different, and that's something that the law is not really aware of.

00:14:09 Brent Mittelstadt

It's been a while since I've met anybody that uses Internet Explorer. I wonder if anybody still uses it.

00:14:15 Sandra Wachter

Fair enough.

00:14:16 Brent Mittelstadt

Yeah, I mean, I guess one thing to add there is just on the translation problem that I mentioned before with standards like you run into the same thing in the context of something like non discrimination.

With that, you know the way that the law has been applied previously is you have regulatory bodies, you have civil society organisations, equality bodies.

That are there to like help bring a complaint or there to essentially you know if you're a judge, you're deciding what equality actually means in practise. And the problem is, you know, you can't take every case of potential algorithmic discrimination to court.

You can't have that human input on it. You need something that is able to be quantified in some way or measured, or, you know, computable in some way. And so that's actually led to a real simplification or oversimplification, really. In research and development around bias.

What we have now is essentially equality has been translated into mathematical definitions, fairness, metrics, BI, BI, metrics.

And what we found and what some of our research has shown is that when you try to enforce those things in a simplistic way, so when you take one of these mathematical definitions and say, you know, oh, my system performs less well for a particular patient group. So I'm going to make it perform better for them.

By sort of erring on the side of caution and just making it making decisions a bit arbitrarily so that they get positive outcomes, what you find is that actually.

The way the.

Fairness is achieved by applying these metrics is by.

Levelling down, it's by making the system equally bad for everybody, because you're quite often limited in terms like the resources you have as somebody using an AI system or you're limited in the data that you have available available to you.

The sorts of solutions that you might want to pursue under equality law, which might involve things like, oh, we need more healthcare resources in this particular region, or we need to increase accessibility to some.

Good. You don't have those available to you in the context of AI, so the only approach you can take is to just make the thing equally bad for everybody, and that's really not the point of equality. Law is to make things worse than they otherwise have to be, and so that that's a real problem in terms of, you know, how you actually translate these laws into practise. So it's great that we have laws like.

They're extremely.

Important, but they were written in a time where we weren't imagining the sorts of problems we have today.

00:16:40 Sandra Wachter

Yes. And I think one thing I just wanted to add because it has also been researched that Brent, Chris and I have done over the last couple of years is that we actually did a legal assessment of those fairness tools. So those fairness tests so that the tests that supposed to give you an estimation of how problematic your system actually.

Is and we were assessing it whether it clashes or is compliant with European non discrimination law. And so we looked at the 20 most common biassed tests that are currently in use, and out of those 20.

13 so 2/3 clash with European nondiscrimination law, and so one of the main reason why this is the case is because the vast majority, if not all of them, were developed in the US, where the equality law system is just fundamentally different than it is in Europe.

Europe plus those tests were developed by tech scholars who don't necessarily talk to lawyers and lawyers don't talk to tech people very often, and there is very little collaboration exchange between different countries. And so you end up in this problematic situation where good people want to do the right thing, create biassed tests, and good people want to do the right thing.

Using them and you might be in Europe wanting to use such a test to make sure your system is actually fair and accurate, but you're not aware that you're actually clashing with the law. And so that's really, really problematic. And so I think that just shows why it's really important to have.

Of interdisciplinary collaboration and really, really collaborate with people that are outside of your own legal discipline as well to make sure that you're not creating systems that might be illegal in different country.

00:18:20 Veena McCoole

OK, it sounds like translating standards into practise kind of opens its own kind of worms. If we move on to talk about large language models such as ChatGPT, you've both published research on the potential societal harms of these technologies.

I'm curious what's at stake here and who is being affected.

00:18:40 Brent Mittelstadt

So I'll I'll start with what I think is sort of the most wicked problem of these systems are the the hardest thing to solve. So there's been lots of talk about hallucinations, which is the definition varies, but it's essentially when a language model tells you something incorrect, when it gets things wrong.

When it claims that there's active volcanoes in in England, for example, and I I've checked, there's not active volcanoes in England, despite what Google has told me.

The problem is so you have these hallucinations that are very obviously wrong and those are easy enough to fix. You can fix those by, you know, feeding in better data by having people who annotate the responses and basically say, OK, this is a correct response. This is an incorrect response.

And the things that are obvious and don't require any sort of specific expertise or prior knowledge, easy enough, you can fix those. Still very difficult problem, but you can fix those.

The wicked problem is what we call careless speech. So that's hallucinations that are subtle. They are things where you have to have some prior knowledge to recognise that a language model has told you something incorrect. So if, for example, ChatGPT, you asked about movies Tom Hanks has done and it says, oh, Tom Hanks was in a wonderful movie.

About the Apollo 11.

Well, actually you would have to know that the movie that Tom Hanks was in was about the Apollo 13 mission, not the Apollo 11 mission. In order to recognise that that was incorrect. That's a very simple example of not, you know, particularly high risk. But we've had other examples since these models have been released that show you just kind of.

How poorly understood of a phenomenon this is, and potentially how significant it's societal like impact?

Act is. So there was this great example where there was there used to be a sort of race based science movement and some research that was done within that movement tried to calculate the IQ, the average IQ of people from different countries in the world.

Results were very biassed, and the the research has since been debunked.

A researcher that was familiar with that sort of race based science movement was doing some research that was related to.

To it and searched in Google quickly. You know what is the IQ of? I think it was Ethiopia that they searched for and Google Gemini. So the the AI previews that Google gives you very confidently responded. Oh, you know the average IQ of Ethiopia is this particular number.

And that researcher, because they were familiar already with the debunked research they were able to recognise. Wait a minute. I think that number is coming from that debunked research. So they checked. They did looked at some other countries. They checked multiple search engines and it turns out that what these engines were doing was just directly sighting from this data set that was recognised as racist and.

Debunked. And so you can imagine if that researcher was not doing that particular query at that particular time with that particular prior knowledge, you could have Google giving these racist results to countless numbers of users. That's an example of sort of the why this is such a difficult problem to solve, why it's not just as easy as.

Very obvious hallucinations to solve.

And it doesn't really seem like there's a clear path to a solution for that.

There's this related problem of model collapse, which is when you have these system, these systems require massive amounts of data and labelled, you know high quality data in order to improve over time. But now the data that they have available to them has been sort of polluted by what is called AI slop or basically.

You know, texts that they have generated pictures they've generated, they're now being fed back into the system and used to train the next version of the system. And so you have this prediction that you'll eventually see these models collapse in on themselves or basically no longer improve because they're being trained on their own outputs. And there's been some recognition of that.

From users from sort of the general public, there's even a browser extension now for for Google Chrome I think.

Where it will force if you do a Google search result it will force Google to only show you results either from 2022 and before 2023 and before and sort of in recognition of I want information that's from the pre AI era I want before I want information from before everything was sort of polluted with AI slop and so that that's to me the starting point of much.

Problem where we don't know what the broader impact.

It will be.

00:22:59 Sandra Wachter

Yes. And just to say that Brent's point is really, really important and is really important to understand that this is not an anecdotal problem in the sense that there are statistics that show that.

Generative AI when you ask it something, it's wrong in 60 to 80% of the time. That's a lot like you're almost better flipping a coin at this point, and so that's something that is really, really important to keep in mind that even though it's.

Written in a very convincing way, purposefully written in a convincing way, it is wrong most of the time, and unfortunately the law doesn't really know how to deal with it because it's a type of.

Speech that is usually it regulated like law, cares about speech that is immediately harmful so.

About insulting people.

Discriminating people. I hate speech right? Where you're doing something to hurt somebody. Law cares about speech that is being used to deceive somebody like to defraud, to fraud. Somebody out of money, things like that. But, like, lying about the amount of volcanoes in England is not something that the law had ever to deal with because it's not something that people usually.

People don't lie about the movie, said Tom Hanks has been in right? This is just not something that in here in our nature we have other things we lie about that wouldn't be one of them, but for some reason those subtly kind of trivial lies are now.

00:24:29 Sandra Wachter

Supercharged and then used in areas where details intrude, matter of things, science Think academia. Think journalism. Think medicine. Think criminal justice. The legal profession. Investment decisions.

All of that.

Depends on knowing details.

Proof. And if you don't know what the truth is anymore because it's just subtly wrong, it can have a big, big difference, right? Especially in in areas like medicine and journalism and science and academia. And so this is really, really important. And the law hasn't really dealt with it yet because again.

It doesn't have to. It dealt with law that we've liars but with different type of liars that lie for different reasons, and that thing doesn't have a reason why it lies. It just spits out something that is wrong, subtly wrong, and doesn't necessarily harm somebody immediately, but probably.

Cumulative over time and then to a big.

Extend so that's the one thing. Indeed, the second thing I just wanted to to mention because I think that's also something that is very often not discussed in, in public discourse and really something where you have to care about is the environmental impact of those systems. AI doesn't grow on trees.

Far from it. It needs a massive amount of resources. It needs water, for example, because the data centres the the service in the data centres, they need to be maintained, they need to be cooled.

And just to give you an idea of that, it's takes 360,000 gallons of water clean water every day to cool a medium sized.

Data centre that is a number that I can't even understand anymore. There are estimates that assume that AI service electricity demands can grow to the energy consumption of countries like Argentina and Sweden very, very soon.

Which is also a number that I can really imagine anymore. And just from your daily use. Just think about how often you play around with generative eye. So if you generate one image, yeah, if you use Dali to generate one image that cost as much energy as charging your phone fully.

5 interactions with jet GPT cost a half a litre of water.

So just think about how often you use it for tasks that you could also not do that are not really important and think about that there is a massive cost, a societal cost for all of us regardless of where you sit in the world, that we all have to pay the price for.

The last thing I just want to quickly mention has to do with the the future of of work as well, because those systems are able to create text or images in in a way that.

Could be used.

Partially automate certain tasks of your daily routine, right? If you have a system that is.

Just is faster than you. That is, doesn't get tired, doesn't need to go on holiday, it doesn't get sick, doesn't complain. It's much cheaper to maintain. How safe is your job? Actually, in the future? And we have to be really, really aware what that means for.

The future of work in general.

00:27:34 Veena McCoole

OK, you've both shared numerous societal harms. And yet this technology is so prevalent and growing. So who are the winners and losers of the AI revolution?

00:27:46 Sandra Wachter

Yeah, I think in my opinion the, the, the jury still out on that, I think we really have to ask ourselves this question in in a deeper way.

Think about just your personal life. Think about your work. Think about your leisurely time. Think about your interactions with friends.

Has this been improved by AI in the last couple of years? And to what extent do you have more free time now? Do you have more money in your pocket? I have. Things are gotten cheaper. Do you have more quality time with friends and family? And I'm not quite sure that that actually has happened, right? I know that the lives of a billionaire got much better.

I don't know if the lives of the 8 billion other people really got better over the last couple of years and so thinking about that hype cycle, that always happens with new technologies. We just have to be very, very careful and think about who are the hype man's of those technologies. Is it the 8 billion people?

Or the 8 billionaires.

And hype cycles are very common. Think about voice assistance. They were all the rave not too long ago, it didn't really take off. Driver wise cars were all the hype that didn't really take off crypto blockchain. Same hype cycles didn't take off.

In the end, and so whether or not this will actually last, I think the the jury is still out on it, yeah.

00:29:07 Brent Mittelstadt

I would, I would agree with that. I think we don't know who the we don't know the the scope and extent of the losers of the revolution.

Solution. We know who some of the winners are at this point.

And we're being sold a vision that there will be many more winners to come. I always challenge people to tell me a positive use case of AI when I talk to them and without fail, everybody mentions protein folding.

And there's a couple of examples from medical research that people quite often go back to.

And yet you can come up with countless numbers of other sort of negative examples where it's very clearly beneficial to a company or to the, you know, the owner, the CEO of a company, but to the people whose jobs are being replaced, or at least.

Vastly changed where the things they actually enjoy about the jobs, like codeine or building systems are being taken away and replaced by the things they really hate about the jobs, like gathering requirements, doing customer service, that sort of thing.

I think the vision that we're being sold by these companies is vastly different now than the reality and I agree. I think that's something that that needs to be given much more attention in the context of regulation in particular is if you don't have a clear societal benefit planned to your system, if you can't articulate what exactly the benefit of your system.

For society, I think you should face much harsher requirements and we should really be encouraging truly social.

Beneficial innovation as opposed to just vaguely socially beneficial innovation.

00:30:39 Veena McCoole

OK. So on that point, people are hyping up the societal benefits of AI. They're comparing it to the printing press and the steam engine. Are those comparisons justified?

00:30:50 Brent Mittelstadt

So I I'm.

Always concerned with those comparisons because they tend to be made by somebody who's leading a start up. It feels like they're going for the next round of funding, and so they need to sell this vision of they've created a technology that will be just as disruptive as the steam engine or the.

With the steam engine, we'll see. We'll we'll see whether it truly, really leads to serve, you know, something equivalent to the industrial Revolution, or if it's a flash in the pan, kind of like blockchain was sorry to offend anybody. It's really into blockchain and crypto.

Or driverless cars appear to also be a flash.

The printing press, I think is a really interesting comparison because it feels like generative AI is doing the exact opposite of the printing press.

The big sort of threat to the printing press when it was invented was that it decentralised the ability to spread ideas and spread information. It put that into the hands of the masses. Both put the information into the hand of the masses, but also made it easier for people to to spread their ideas. And that was very threatening to the institutions that previously controlled that ability.

The only ones able to spread ideas at scale.

And now we have generative AI where, you know now you've had the Internet where anybody can post anything they want. You can consume whatever you want, seemingly the whole vast scope of of knowledge that humans have generated over time is available to you at at your fingertips.

And we're returning to a time when suddenly the information is disseminated by a handful of companies. Suddenly rather than gain search results, you get an AI preview that tells you everything that you need to know. That answers your question directly, and in doing so, steals the information from somebody else who creates.

00:32:28 Brent Mittelstadt

I did it so it feels like. Where's the printing press decentralised that power? Suddenly generative AI is very strongly centralising it again in a way that's really dangerous.

00:32:40 Sandra Wachter

Yeah. And I I also think that we we we pay much less attention to.

Whether we should be automating all types of tasks just because we can, which seems to be the thing that we currently do, is just.

If you can put AI into something, definitely do it. And so I just think what that would mean for my own life. So like, I'm I'm an academic. I love reading. I love writing. I love thinking. The idea that this is outsourced to an AI system makes me really sad and angry.

Because that's the the whole reason why I'm in this game. So the idea that.

That I can spend. I don't. I don't quote unquote have to do this anymore. Seems like a step back when in reality, what I would like to see is that I have a life where I can spend much more time on reading and writing and thinking.

Rather than have AI take those things away from me.

00:33:38 Veena McCoole

Got it. Brent, you're a philosopher as well as a data ethicist. What philosophical questions does AI and its regulation raise?

00:33:48 Brent Mittelstadt

Be a lot of the some of the the questions I've alluded to about like aligning AI with, you know, social values with the sort of foundations that society has been built on, and in particular, about what is the future, what is, when we talk about ethical AI, for example.

What is the good life, the future that we are trying to bring about by developing AI and using it everywhere?

I don't think we talk about that explicitly enough, and often enough. I think we have a vision being sold by companies, you know, by venture capitalists about all the societal benefits that AI will bring about. As we've already talked about, whether those have been realised is highly questionable.

But I think we should actually say, OK, but if you completely get your way, if you have no regulation, if you have infinite money, you can develop this technology and use it however you want. If you can bring about artificial general intelligence like some people believe they can do or want to do.

What exactly will that future look like and who will it be good for?

Because the the implementation that we're seeing now of AI seems to be one that divides people, that it's clearly beneficial for very few number of people and harmful or at least neutral for the vast majority of people.

And so, you know, if Elon Musk gets his way, if Mark Zuckerberg gets his way, what is this future that they want to bring about? If you start to delve into some of the sort of schools of thought or like the the base philosophies that are pushing a lot of innovation and especially in silicon?

Valley they believe some really weird stuff and some stuff that is really worrying. If they were to actually fully get their way, we don't have time to go into all the sort of weirdness that comes around with some things like effective altruism and accelerationism. But I think we should be questioning that much more than we currently do. We shouldn't sort of hide behind this.

This myth of innovation always being inherently good and say, OK, but what is your vision of the future and why should I buy into it?

00:35:46 Veena McCoole

OK, this has been a really interesting conversation and there are so many nuances and complexities to how we imagine what a good life looks like in the context of AI development. I'm curious about whether there's research findings or information that still needs to happen to fill the gaps in our current knowledge to help us better understand how to regulate AI for the long term.

00:36:08 Sandra Wachter

Yeah, I think we're really at the point where we have enough already to do something. I think for the longest time, regulation has been halted and curb because somebody was like, no, we need more information, we need more evidence. No, no, no. We have evidence for decades now.

If not more than that, it's really time to act and to stop pretending that.

Isn't inherently unrisky technology. It is incredibly risky, and regulation needs to step in on crypto riskiness. AI is not this wonderful, perfect being entity that was born, that can only do good. It is an inherently.

Risky, problematic technology that can do good if you design it in a right way, and if you put the right guard rails on it. If you don't do anything about it, it will always be biassed. It will always lead to privacy invasion. It will always be opaque.

It will cause environmental harm. That's just what it is. That is what it's quote unquote, how it was born, right?

00:37:10 Sandra Wachter

And So what you need to do is put some guard rails and regulation around it to curb the thing that it inherently has. So there's enough evidence in evidence gathering is over. It's time to act now.

00:37:23 Brent Mittelstadt

Yeah. The only thing I would add there is I think we need to understand the long term impact of these especially generative AI.

On people's, you know, ability to breed on their learning on their creativity. I think we're rushing to implement these technologies at scale that's.

Are presenting shortcuts for a lot of really fundamental skills. Things that we have traditionally valued, and it's not to say that all innovation's bad. I'm not going to be the person that says that the calculator was a terrible innovation, but at the very least, given how.

Sort of different. These technologies feel how suddenly we are able to replace some really high level sort of types of human skills or human expertise. I think we need to put a lot more effort into understanding the long term impact that has, you know, especially on kids that are growing up with this now and using ChatGPT to do school work for example.

00:38:15 Veena McCoole

Hmm. So with all of those considerations in mind, if you could wave a magic wand and change one thing about AI regulation, what would it be and why Sandra? Let's start with.

00:38:26 Sandra Wachter

You. Yes, I think I wanna give a big, big shout out to Professor Barbara Prainsack at the University of Vienna.

Who has done incredible research on data solidarity for for quite some time with her team and and many others have.

Umm. Worked in that area as well. Like Lynette Taylor, for example. And those brilliant thinkers, they they really ask for something that I think has not arrived at the forefront of policymaking yet. But it really, really should.

They argue that innovation should be guided by the idea of public value. So if you wanna get your hands on dicta, if you wanna create an AI system.

You should the regulation that comes from that should be.

Informed by how publicly valuable The thing is that you actually after. So if you building something that is public value for society, then you should be encouraged. It should be facilitated, should be easy for you to do so. However, if there is no.

Public value or very little. Then it should be much more stringent regulation around that and that really shows that we're moving away from this idea of just because we can, we should.

And rather think about should we actually do it and how is it gonna improve our our lives in society? I always use this example of the the London tube. When I was sitting in, in, in in one of the carriages and there was an advertisement for.

Electric toothbrushes that are now also have AI embedded.

And I was like, who is this for? Like, how is the the the teeth clean experience improved by having AI in there? Like show me how this is useful for anything but it cost a lot of money and it cost a lot of resources and it's not necessarily beneficial. So having this idea of public value front and centre for research and innovation.

00:40:25 Sandra Wachter

I think would really tip the scales to make sure that we have AI that is beneficial for the majority for the majority of the population, not for a couple of very wealthy individuals.

00:40:37 Brent Mittelstadt

And for me, I would say the thing I would like to change is to make AI companies actually pay for the data that they're using to build these systems. I mean, we're talking about systems that are built on content that has been generated by people.

You know, over decades, over centuries and now being used in violation, potentially I should say, potentially in violation of copyrights without permission.

And there are many legal cases that are going forward now to see whether we can make AI companies actually pay for this data, or at least, you know, pay for their violation of copyright. I would like that to be a settled thing. I would like it to be clear that they do have to pay for that, that, you know, the the artists, the authors, the painters, the voice actors.

Whose content is essentially being used to build and improve these systems, that they actually get their fair, you know, cut of it. And they have a say in how these systems are are developed and used, because what we've witnessed so far might be, you know, the largest.

Violation of copyright and theft of intellectual property that the world has ever seen.

00:41:38 Veena McCoole

Those are both great wishes. And finally, as we wrap up what is something we should all be aware of and on the lookout for in terms of AI developments in 2025.

00:41:49 Sandra Wachter

I think what we should be aware of is that we really are witness.

How history unfolds itself. We have so much new regulation that is coming into force and it's going to be enforced in the next couple of years. This is really like the point where the rubber hits the roads. I'm super excited.

And terrified to see how how that plays out because the last couple of years have been really reproductive in terms of standard setting and regulation. And I'm super excited to see how and how effective this actually is and hopefully this will lead to us having better and safer products on the market.

00:42:25 Brent Mittelstadt

So I think we'll see much more development and usage of what are called AI agents. So essentially just AI that will do things on your behalf autonomously.

Honestly, that's been predicted to be the next big thing in in 2025 and I think we'll also see some claims that people have created artificial general intelligence. And as always, I think we should be very sceptical those claims cause at the very least, the definitions we have, the benchmarks we have to say that something like that has happened, are very flawed and if we follow.

Sam Altman. Then it would be any system that makes you know X amount of money over time would be artificial and general intelligence, which is utter nonsense.

00:43:05 Veena McCoole

Sandra and Brent, thank you so much for your time today. It's been fascinating to explore the world of AI regulation with you. And thanks to our listeners for tuning into this episode of the OII podcast, if you've enjoyed it, please leave us a review and share the link to this episode with your network.

00:43:20 Veena McCoole

We'll be back soon with more conversations about the digital world. Take care.