AI and You

Transcript

Guest: Ryan Abbott, part 2

Episode 51

First Aired: Monday, June 7, 2021

Hello, and welcome to episode 51! We're going to conclude the interview with Ryan Abbott. He is Professor of Law and Health Sciences at the University of Surrey in the United Kingdom, and Adjunct Assistant Professor of Medicine at the David Geffen School of Medicine at UCLA. His research has been featured in the New York Times, Wall Street Journal, and Financial Times, he has published widely on issues associated with life sciences and intellectual property in leading legal, medical, and scientific books and journals, and *Managing Intellectual Property* magazine named him as one of the fifty most influential people in intellectual property in 2019.

We are talking about issues raised in his 2020 book *The Reasonable Robot: Artificial Intelligence and the Law*, and last week we talked about the possibility of AI owning patents or intellectual property, and law and regulation around self-driving vehicles. This week we're going to wrap up talking about self-driving cars, liability and punishment for AI infractions, and rationales for changing our taxation system for AI-generated wealth. Let's get back into the interview.

Where does *intent* play into the issues of liability and our perception of AI, and any culpability that it may have? As a species, we tend to judge humans by their intentions and machines by their actions, because we don't think machines can have intentions and that may or may not change as AI gets more sophisticated; but it is a rather important distinction, and where do you see that distinction becoming important or changing?

Another very interesting concept, it depends a bit too also on the area of the law. So, for example, with civil liability, let's say I get run over by a self-driving car; sometimes we do but generally we don't care that much about intent. The question is would a reasonable person have run me over, Yes or no, liable or not liable? And that makes sense because what I usually want out of tort law is to have a system of law that results in fewer accidents. And I often really don't care what someone is doing or why they cause the harm or why they ran me over and I also don't really care if an AI with a logic-based system or an expert system or an unpredictable neural network ran me over, I just really don't want to be run over. And so by focusing on behavior, it kind of puts aside what people meant to do and it gets at what they did do and it encourages them to act in ways that will not cause harm. That's not the case with every area of the law and I think the area where it's kind of least the case is criminal law. So, running someone over may also be a crime and there we very much care about someone's intent that you wanted to run me over and at least that your intent was so bad, you just were reckless with whether or not you would run someone over and that is a very interesting thing to think about machines doing because there we want almost to punish people in virtue of things like culpability and retribution. And it isn't clear it ever makes sense to think about a machine having an intent or being culpable or retaliating against the machine because indeed machines don't think the way that people do. On the other

hand, in the book one of the things I get into is, almost more of a philosophical thought experiment about whether you could hold a machine liable for a crime and whether it would make any sense. Legally of course you could change the law to do that. There are jurisdictions that attribute legal personality to all sorts of things. New Zealand recently gave legal personality to rivers. There are some countries that have looked at having some legal personality for a machine, so it's almost less an issue of *could* the law do that then *should* the law do that? It is also the case that we already hold artificial persons criminally liable for things, mainly corporations. There are different views on that. You can also hold corporations liable without intent for strict liability crimes, even sometimes for failing to act where a corporation had a duty to act. So, holding an artificial person liable for failing to act without a mental state is something our law already accommodates. And criminally, punishing the machine could have some benefits. So, for example, if a self-driving car started running over law professors, we could convict the car and destroy it and say to law professors everywhere, we will not tolerate the sort of conduct whether it is coming from a person or a machine. If Tesla made that car, it could deprive Tesla of its valuable AI and thus change Tesla's behavior, right? As a practical matter, probably no reason ever to do that; it would be a dramatic legal change. It could have troubling outcomes too and that people start thinking of machines as being morally on par with people as agents to the extent that really what we would be trying to do with that criminal law is changing the behavior of the people making and using and developing AI. Probably better to go after those people directly with civil liability rather than criminal ones, or criminal liability where they have directly had some sort of harmful intent. So, what I think through in the book is basically this is something the law could accommodate. It is probably not a good idea for any kind of reason but it isn't really something that's that far out and depending on how you think of the nature of intent, is something that you could potentially impute to a machine, just wouldn't be a good idea to do it.

It seems that runs into the question of the origin of volition. A corporation - and corporations have been likened to superintelligences - a corporation is a container for people, so if you punish the corporation you are punishing the people inside it, they have motivation to act on that. It seems the point of punishing the corporation is that those are the people who will cause the corporation to act better. An AI - absent say the level of AI that's now evolved to human-like capabilities of thinking, intelligence and consciousness and free will that we will leave off the table as you did in your book – is not – you could punish it as much as you want and it creates no reason for it to change its actions unless those punishments flow through to some*one* who is directing it. Is that not the case?

Well, that is the case. So, but let me talk about that and then talk about the corporation bit, so in criminal law we would distinguish this between specific deterrence and general deterrence. So, let's say that I ran someone over with a car with an intent to kill them and I got sent to prison. One thing that would do with it would deter me from doing the same sort of thing again were I ever to be in a position to do it again; that would be specific deterrence. The other sort of thing it would do would be general deterrence, which would be to everyone else who sees me going to prison for this saying, if I am ever tempted to do that, I am sure not going to do it. Now, for sure,

unless you designed an AI to take account of criminal law, it couldn't care less if it goes to jail or other AIs go to jail, right? But if there was a punishment for an AI behavior that had a corresponding negative outcome on someone with a financial interest in it, they would be deterred from doing it. So, for example, if every time Tesla's self-driving car turn light came on the wrong way, that Tesla got criminally convicted of a traffic infraction and the owner of it had to pay for the ticket, right? At the end of the day, the people you are impacting would be the people making, building and using the AI. And indeed, this is one reason why even though the law might focus on AI behavior, it has to be designed in such a way that you are really looking to impact the behavior of those people. Corporations are interesting, and there are kind of two veins of thought on this. One is that corporations are just a some of their agents, and there's really no need to punish a corporation. There's always someone at the corporation who did some bad thing and they are really the person you should be going after. In fact, punishing a corporation causes harm to a lot of innocent people, mainly the shareholders are the ones who are punished when a corporation is punished. And there's a hope that if a corporation that you own stock in does something that results in a criminal punishment and a fine, you will punish management in turn by getting a new management team or assisting on accountability. But there's another theory that is more kind of what I think you were getting at earlier which is that a corporation, while composed of people is something more than the sum of its agents; that people do individual things that not necessarily criminally culpable kind of individually but in the aggregate do something beyond any of them. Or there is groupthink involved in companies, where people act in synergistic sorts of ways and really for thinking about it in terms of punishment, the corporation is something distinct from its agents that has done something above and beyond what they can do and if you think about that in the AI mold, I have a lot of sympathy for that view of things because an AI really isn't necessarily just the sum of things that people have programmed in it. If you have open-source program, you could have hundreds or thousands of people programming bits of it and very hard to disaggregate who has done what or contributions made by different people may combine in synergistic ways that make it an entity in its own right beyond that which the programmers made. And similarly where you have distributed ledgers, there may be need for not reducing behaviors by distributed entities to those of individual people because it doesn't line up very well.

I am just curious, are there any non-computer technology analogs of this in legal precedent of things that were created by so many people that you couldn't trace the accountability back to any one person or even group of people reasonably, but there was a consequence from what they created, that had some effect on someone?

Right. Well, the first thought that sprang to mind was, well this is indeed very much the whole idea of the corporate criminal liability, which is either because theoretically we think corporations are more than a sum of their agents or practically because for various reasons, we cannot go after individual corporate agents, we do think of a corporation as more than the sum of people and that it is effectively doing things and that it is a target for punishment. As to whether people have set something in motion that goes on to cause harm. Well, yes.

Well, how about climate change? You have industrialized companies putting carbon into the air and that causes islands in the Pacific to become covered in water and their people are displaced. Someone, a group of people, created that situation but you cannot trace it back to anything you can sue.

Right, so irreducibility and unaccountability are definitely challenges for the law. There are definitely harms that are caused where you cannot identify a particular actor from that. I was also thinking a bit about animal-caused harms and there are various people like Kate Darling, who talk about AI a lot through the lens of animal law because animals have somewhat of their own agency, are owned by people often but not always can cause harm. Sometimes, this has to do with the way that a person has kept or trained a dangerous animal, sometimes not though, and that certainly has some machine analogs. The climate change one is a very complex one, governed by a series of international agreements on climate change and also by national. and in the US state regulations on what one can do with the environment. Sometimes, there is definitely liability for environmental pollution and sometimes you are allowed to pollute to a certain degree. This may cause a harm and you are not liable for it and if you are an island that is getting swallowed up by the sea there's very little legal recourse for you. So, from an accountability perspective, I definitely agree that's a major challenge analogous to maybe some harms caused by distributed systems. If you look at a distributed AI, so something on a block chain that no one is controlling, to find liability for that, you either have to find someone who owns the AI or who developed it or designed it or implemented it and that they did so in a way that was legally wrongful, because they were negligent in how they designed the thing or maybe setting it into motion in the first place. But ultimately what you are looking at is that the person did some wrong thing in relation to the AI, not necessarily imputing the AI's behavior to the person.

I want to shift in our time left to the impact of automation on personal income because you devote some space to the consequences of automation, increasing inequity, the way that capital flows in that and some of the work that has been done and suggestions floated about universal basic income and others. First of all, are you looking at that through the lens of law or macroeconomics?

I am dabbling a bit in economics, which I shouldn't be doing, but I am mainly looking at sort of the sorts of legal solutions we could bring to bear in response to this. So, we talked about how if I can get an AI to replace a person at my company, I can save money on taxes by doing that, right? And that's problem one, not that we don't want to automate where it's efficient but we probably don't want businesses doing it just to save on taxes. Problem two with this is that when you do something like that, the government loses tax revenue and it does this for a couple of reasons. (1) It does depend on the country - but let's use the US - and say you lose out on payroll taxes that the company would have paid for the person. (2) That person then no longer pays taxes based on personal income, which are taxes comparatively a very high rate compared to corporate taxes and even if the corporation becomes more profitable through that automation, which they may not a relatively small amount of that is remitted to taxes, both from statutory corporate tax rates and effective corporate tax rates being substantially less than that. So, what you get with automation is companies potentially automating to save on taxes, paying less tax revenue, and

what you end up with is the government having less revenue with more people potentially out of work. That's problematic for at least two reasons. One is that generally the accepted solution to "machines are going to put people out of work" is where we are going to retrain them in new job types and provide some social benefits for them while they transition. But for the government to do those things, they need more money not less and so automation may actually be reducing tax revenue. And the other is that AI is likely to generate a massive amount of wealth, probably poorly distributed and this is going to contribute to distributional problems and wealth of the United States. Now, how that plays out exactly will depend on the future of work and how good AI gets at automating jobs versus augmenting jobs and how quickly it does that sort of thing. If AI does in fact only put truck drivers out of work in the next two decades and all these truck drivers go on to find new better paying jobs, well it does work out better for everyone in the end and that has been the history of the automation challenge in the workforce, though even there historically as a society, we haven't done too good a job at transitioning people into new jobs and helping the people who are really kind of getting the short end of the stick with automation. So, I think we could do a lot more with that. But the other vision of the future of work is that AI gets a lot better very quickly at doing a lot of different things, and we are not able to quickly transition people into new job types potentially even large segments of the workforce may be automated. If that does happen, it would be a real problem if we didn't have a solution for dealing with that and that sort of thing probably is universal basic income, whether or not we should have that sort of thing now. And if you are going to have that, we should be taxing some of the wealth being generated by AI. Now, how would you do something like that? Well, you could tax AI literally, like have a robot tax, but there's a lot of problems with that. (a) it punishes businesses that might legitimately be more efficient with automation, hard to define a robot for tax purposes, robots and people may not be one-to-one interchangeable, there would be a lot of gamesmanship with the IRS, and so what I am effectively suggesting is we indirectly tax AI by increasing capital taxes and decreasing labor taxes. So, for example, you get rid of payroll taxes and thus we're not punishing businesses for hiring people anymore and how are we making that tax up? Well by increasing taxes on companies, right, although there's other ways you could make those taxes up, but essentially, when you have AI which is really in a sense capital itself taking over the role of labor, I think we are going to need a more neutral system between robot workers and human workers.

I think this is the big conundrum of automation in that when we project or fantasize about the future with AI and automation and robots doing things for us, making life easier, we see that as being equitable. We see that as being fair and lifting everyone up; but in reality, the people who are put out of work by automation tend to be, as you say, the lower rungs, and until and unless automation puts the CEOs and the lawyers out of business, it is going to be the truck drivers and the burger flippers. And so that has this result of transferring capital from people who need it to people who don't. So, what you suggest there, that form of taxation is, I think that's the real interesting part because to get to the future we want, that's fair and equitable, we somehow have to cross this divide of "how do we actually take the money from one place and put it in another" according to the system we have now, without someone screaming and

voting you out of office. And is your proposal tantamount to saying we are going to tax a corporation based on the ratio of its net revenue to its employee base?

Well firstly, I thought that was a great response and again, these are not entirely new problems and I think also kind of the reason why you don't want the law to just wait for 20 years, see what happens and then try and figure out a solution to it, we want to make sure that with taxes these are done for various reasons including some amount of distributable fairness or distributional fairness, right, that we have the right system in place before this happens and mine kind of addresses, I think, either the scenario where we just get rid of all the truck drivers or where also we get rid of all the lawyers, whether or not just a lot, you know, a whole lot or a few people are automated. One of the things the book does is say, if you buy into this idea that we should have a more neutral tax system, how would you do it? And there's a lot of ways you could do it. You could tax robots, you could have like an unemployment by technology system where employers pay into that and you have something like unemployment benefits, You could have greater taxes on industries that work with fewer employees. So, if you look at kind of Google versus Ford Motor in the '80s, you know, much higher revenues, many fewer employees, right? So, you could kind of preferentially tax companies that use automation. I don't think, personally though, I think the best way to do it is to do it in almost the most neutral way possible so that you are not punishing legitimately effective businesses and you are not encouraging gamesmanship, which is simply that the tax burden on capital should be higher than it is now and the tax burden on labor should be lower than it is now. And I think firstly, that frankly that should already be the case, but AI is going to exacerbate that divide and so it gives us a new impetus to address an existing problem now before it gets worse.

Well, that is so thought provoking and I am just conscious that our time is drawing to a close, and what frustrates me is that there's so little time to discuss this when this is a conversation I think that should be being heard by thousands of people that are not just trying to fit in a one-hour interview, but that it is shaping policy and it - well, I am just venting here - but I wish I saw more evidence that this was being taken seriously. Is it fair to say that your book is an attempt to push the boundaries of this, to point out to people, "Here's where we are going, do you like this or not, do you want to do something about it?" Do you have an agenda like that?

Yes, definitely. I mean, this book is both an attempt to help people think through some of the challenges and opportunities that AI is going to pose us as a society and how the law could deal with that in ways that provide the most benefits.

Well, fantastic. The Reasonable Robot. We will have a link to that in the show notes and more information about that in the notes after this interview concludes, but how can people keep up with whatever else you are doing and follow you or get in touch with you as they might be inclined to?

Oh, well, I would be delighted. I am not terribly hard to find online. So, I am on Twitter at DrRyanAbbott, though I don't tweet that much; and on LinkedIn, and would love to hear from viewers.

Terrific. Do you have any projects coming down the pike you want to mention?

Well, I think those patent test cases are kind of getting some good attention so we are very excited about that. The UK Court of Appeal has accepted an appeal from the High Court on a discretionary basis. That hearing is in summer and they are debating whether or not you can have patent protection without a traditional human inventor. That appeal is before the European patent office board of appeal. The European patent office president has just filed an intervention in that to make a comment on it. There's a hearing at the US district court next week, I am participating in that and the UK action. So, we have been very excited to see mainly the IP community take an interest in this in courts, take an interest in what technology should, what IP law should do to deal with advances in technology and I think that thinking like this is warranted in a whole lot of areas.

Well, thank you. And would you let me know when you have an outcome in that or some progress so that I can keep up with what's happening in those cases?

Absolutely. And if anyone's interested in it, I made a short website artificialinventor.com, where there's updates but I will definitely keep you posted.

It is absolutely fascinating. Ryan Abbott, thank you very much for coming on the show.

Thanks so much for having me.

That's the end of the interview.

I am struck by how far hitherto science fictional issues have now made it into the mainstream. The idea of how to restructure the economy and our system of wealth distribution is now reaching levels that were once only discussed by very marginalized people or in very esoteric context, but now are being given serious consideration by people who are closer and closer to the control points of those systems in our world. I also really appreciated hearing these questions of criminal culpability and civil liability dissected with a lawyer's precision. It shows how people are starting to realize the scale of the impact of AI that we're heading towards. Is our society going to be ready for these changes by the time they arrive? What do you think?

Apropos of this entire discussion, listener Paul writes that, "neither Free Will nor determinism have been proven and that therefore either one is only a belief. We might be playing the game of Life by the wrong rules and not know it." Our system of law pivots quite significantly between free will versus determinism on questions of in what circumstances someone can be held criminally responsible for their actions, and doesn't that also come into play when an AI commits some offense, as to whether it was deterministic and where the responsibility lies?

In today's news ripped from the headlines about AI, and since we were talking about self-driving cars, the California Department of Motor Vehicles has granted the company Nuro a permit to deploy autonomous vehicles on public streets in two counties near the Bay Area, starting with Prius vehicles in fully autonomous mode, followed by their custom-designed electric R2 vehicles. They have tested the R2 vehicles in three states with no drivers, occupants, or chase cars, which is pretty gutsy, since even the Waymo service in Phoenix operates with roadside assistance cars within 5 minutes of their self-driving taxis. Nuro plans to use their vehicles also for automated deliveries.

In another headline that tells you the whole story and falls into the category of shocking-to-no-one, I quote, “NYPD to dump robotic police dog following backlash.” Yep, the NYPD had obtained a Boston Dynamics quadruped robot from their Spot model line, which they called “Digidog.” It went on patrol with officers. You can find videos on Twitter with quite not-safe-for-work language from the bystanders. After a storm of outrage, Digidog has been retired. So, for the benefit of other police departments shopping for mechanical mutts, where did this go wrong? Let us count the ways.

First, Boston Dynamics expends zero effort making their robots look even vaguely cute. Their designs, which, don’t get me wrong, are highly capable, are purely functional, with the side effect of looking like something dredged from a robot nightmare. If you’ve seen Robocop and you remember the seen with the ED-209, this comes off as a smaller version of that. Second, the name “Digidog” isn’t winning any hearts or minds, and painting it on the side in block letters didn’t help. At least try something like Rex, or Prince. Third, debuting it in a housing project sends exactly the worst message – I mean, do I have to spell it out? If you made a heatmap of the NYPD’s reputation around the city, that place would be the coldest spot. If on the other hand you wanted to show that you found a new way to express your attitude towards the occupants with even less humanity than previously thought possible, well then, mission accomplished.

Now, look. There is absolutely a place for robots in the police department – get *Robocop* out of your mind for now – in roles like bomb disposal, hostage situations, hazardous materials, or dangerous environments. We want to see those use cases succeed. Let’s also be clear that Boston Dynamics said they would not allow their robots to be armed, although that’s nowhere near enough to tilt the publicity war in the NYPD’s favor. Make some mods to the head of the robot so it looks at least somewhat canine, with features that suggest eyes, snout, mouth, ears. Show it walking around upper-income neighborhoods being petted or fetching a stick. Especially, get a video of it being used in a situation that’s too dangerous for a human officer. If the robot gets damaged or destroyed in the process, even better. This ain’t rocket science.

Next week will be a very special two-parter, because it will be our 52nd episode and you know what that means, it’s our one-year anniversary! And for a lighter episode, I will be talking about AI in fiction: books, stories, TV shows, movies, as I mentioned a few episodes ago. That episode is coming up next week, but I won’t be doing it myself, I’ll have two very dear and talented friends with me: Doctor Robert James, who is a published expert on the Academy Awards, and Jim Gifford, who is the bibliographer for the science fiction author Robert Heinlein, and he was the excuse for us all to work together in 2007 to produce the Heinlein Centennial, a science fiction convention unlike any before or since. We’re going to have a lot of fun with this and you’re invited to join us. That’s next week on *AI and You.*

Until then, remember: no matter how much computers learn how to do, it’s how we come together as *humans* that matters.

http://aiandyou.net