

AI and You

Transcript

Guest: Ted Parson

Episode 30

First Aired: Monday, January 11, 2020

Hello, and welcome to episode 30! In this week's episode, I'm talking with Dr. Edward Parson of the AI-PULSE project at UCLA. That puts him at the intersection of artificial intelligence and law, on the leading edge of research in the field. Ted is their Dan and Rae Emmett Professor of Environmental Law, and Faculty Co-Director of the Emmett Institute on Climate Change and the Environment. He's been published in *Science*, *Nature*, and many other journals, and his recent books include *A Subtle Balance: Evidence, Expertise, and Democracy in Public Policy and Governance*. He spent twelve years on the faculty of Harvard's Kennedy School of Government, has consulted for the White House Office of Science and Technology Policy, and has degrees in physics and management science and a PhD in public policy from Harvard. PULSE is the Program on Understanding Law, Science, and Evidence, at UCLA's School of Law. They conduct "interdisciplinary research and innovative programming to study how technological advances and scientific knowledge and uncertainties influence law and policy making, and how their impacts can be managed to advance human and societal well-being." And they've had some fascinating webinars on topics like the balkanization of social worlds into diverging points of view, and alternative payment models for the Internet. Sounds right up the street of this podcast. Here we go with part 1 of the interview.

Ted Parson, welcome to the show. It's a pleasure to have you on.

Thank you, Peter. It's nice of you to have me.

Thank you. Now, we're going to talk about your involvement with the AI PULSE program at UCLA but let's get to know you a little bit first, and maybe you can tell us some of your background and how you got involved in this work in particular.

Sure, happy to. So I'm half-trained in four fields so I don't quite fit in any disciplinary unit. I now have the pleasure and opportunity of sitting in law school, but I am a law professor who's never been to law school. The main theme of most of my work has been on energy, environment and climate policy, law and regulatory issues with a fairly strong engagement in scientific and technological issues. So I have an enthusiasm and sort of comfort with engaging in scientific and tech issues. For the past 10 years, much of my work in the climate space has been on the government's challenges associated with engineering interventions to actively modify the environment, so-called climate engineering, climate intervention, or geoengineering. About four years ago, I got interested in the rapid developments that were underway in the broad space of data, information technology, AI and machine learning, and started up a project investigating those hoping for there being significant parallels or complementarities across the different domains. So I have a joke that now I work on two of the three areas of technology that terrify people - geoengineering, and AI. And all I need to do is establish a project on gene editing and

synthetic biology, and then I would have the trifecta. Now, of course, I can't do that, because in fact, although there's a lot of complementarities, in fact, there are also differences between these two fields. And so one cannot, without limit, proliferate academic or policy work in different areas of technology without running into capacity constraints.

All you have to do is rename it "eschatology studies" and you're in.

Mm-hmm.

Now, the [AI PULSE](#) program; PULSE stands for Program on Understanding Law, Science and Evidence, and now that's got this AI aspect to it. Covers what on first blush are some very broad issues, a broad range. You've had a webinar on it, called the "Echo Chamber," the notion that we're living in social worlds, and that an increasingly narrow, tightly connected and homogeneous - that is the Facebook problem. And then could AI drive transformative social progress? What would that require? And then a thought experiment about whether AI could run the economy better than markets. And each one of those sounds like an entire discipline. Is it just fun to range over so much territory?

Well, it is a great deal of fun, of course. There is an integrating theme here. So when we started AI PULSE a few years ago, our sense was that the debate on the societal regulatory and legal implications of AI and related technologies had gone through a big swing over time. So in the early days, it was dominated by rather breathless, popular accounts about the transformative impact of fully autonomous AI, superintelligent AI. The theme being that these aspects - and it's not that they're to be dismissed - were so transformative, that basically, they were unprecedented in human history and it was necessary to kind of find fundamentally different ways of thinking about those risks. Now, after a few years of that being the main focus of inquiry, there was a reaction led by a couple of scholars who wrote some very sort of forceful and very compelling pieces, saying, "Everybody's talking about the singularity when AI gets smarter than people and in fact, there's all this stuff happening right now that is of very serious legal and policy import and people should be paying attention to that." So the center of mass of the debate shifted from the distant future to immediate issues where AI, big data's surveillance machine learning, were impacting present legal and policy controversies and concerns in a very direct way. So the bulk of the debate shifted toward things like racial bias and algorithmic decision making, transparency and accountability, and what those legal and policy requirements implied in the domain where decision making was increasingly automated and algorithmic in character. Now, when we established AI PULSE, sort of the basic insight that animated it was that there's a big gap in between these two. There are absolutely immediate societal impacts and regulatory and policy and legal concerns posed by stuff that people are doing with AI right now. I'll keep saying AI even though I should, at every point, put an asterisk on that, say, "AI and related data and technologies" because there's a cluster of sort of neighboring technologies and capabilities that have complementarities, and they extend into other space. Okay, so superintelligence, singularity, remote future shifted to compelling immediate concerns right in front of us, and then this space in the middle relatively unoccupied. So we have tended to focus our interest on that space, attempting to characterize potential applications of AI and related technologies and data

that are, say, five years or 10 years in the future, and bring what relevant knowledge is available together to characterize the nature of the impacts that might be expected from that, and thus the nature of the legal and policy challenges that they might pose. So at that level of generality, that kind of ties together the activities that we've done in AI PULSE. So the piece you referred to on the thought experiment - could AI run a centrally planned economy, not just better than the failed Soviet experiments of the mid-20th century, but also better than markets - that was an attempt to identify a potential application enabled by the vast expansion of capabilities that's now underway and that would be transformative in its impacts for good or ill or no doubt some complicated combination of good and ill. The project with my colleagues at the RAND Corporation on the potentially transformative impacts of AI for social good, that's another direction of efforts to look 10 years ahead, what might human welfare gain from these technologies? What would be the requirements or conditions for having a chance of gaining those? And how, if at all, do they look different from what's happening right now? The Echo Chamber and the other webinar we did on payment models for internet content and their implications, what we're doing there is trying to identify current concerns that are much bigger in their potential impact and implications over a decadal scale future. So the Echo Chamber, basically the claim that social media and in particular, the algorithmic enabled feed up of information and connections via social media is fragmenting social connections, bringing people into more tightly connected, smaller bubbles. So if that is correct, about the phenomena and its causation, it has huge implications for social, economic and political life over a medium-term time horizon. So you also might have noticed that in order to shake up our thinking on that, we actually invited in a brilliant science fiction writer to be among the participants in the conversation. And so, anytime we were tempted to go small bore and immediate - we legal scholars - David Brin was there to say, "Well, but you could think of this in bigger context."

Right. I listened to that. And so there is this, perhaps tension in that you're looking for things that are deliberately five to 10 years out. On the other hand, the question of social media and its effects upon society and democracy are current right now. I just finished reading Brittany Kaiser's book about Cambridge Analytica and what it did for the 2016 presidential election and the Brexit vote and other national votes, and that is the application of AI towards psychological warfare of a kind - propaganda. What are your thoughts about the immediacy of issues there? How has your thinking developed about what kind of regulatory response might be needed?

Well, the most familiar trope about legal and regulatory responses to disruptive technology is that law and regulation are too slow, too clumsy, too rigid, and too behind the curve in terms of scientific and technical expertise and so the whole endeavor is kind of impossible. Now, in my view, that council of despair is strongly overstated, and yet, there is some truth to it. I mean, challenges and disruptions from technological advance come up, sometimes very suddenly but not always as suddenly, as startups say they're going to when they're pitching their enterprises to their early investors. But I think some of the things that are happening now in terms of deployments and applications of AI, big data and related to tech, speak very precisely to current fault lines, concerns and controversies in society. And if you project the impacts of them forward, you don't get big kinds of phase changes in what you expect. So think about something like algorithmic bias and decision-making. It's now very, very well established that machine

learning based systems, because of their needs to be trained upon large volumes of historical data, tend to embed the habits of decision including biases and prejudices that are embedded in the aggregate history of human decision-making. So criminal justice - risk assessment algorithms to be used in parole, bail, sentencing, etc. - now, it shouldn't have come as a shock, but it did initially come as a shock, that these systems had very clear racial biases. In that respect, they are sort of resonant with racial biases that exist throughout current human society, American society, and broader society and they represent a challenge where the introduction of automated decision-making via AI changes things. It doesn't necessarily make them worse. I mean, I think the best characterization of what happens with AI is that the systems are less prejudiced than the worst current human decision-making institutions, more prejudiced than the best, and might actually have the constructive effect of surfacing for examination, biases that were previously concealed. But I don't think that the structure of those challenges and potential responses is likely to be really different in five or 10 years than it is now. And that last point is the distinction I'm trying to draw here. Now, if we compare that to say something like highly skilled manipulation of individual decision making through machine learning driven algorithms, optimizing for engagement and retention, we absolutely do have debacles already such as Cambridge Analytica's role in the 2016 US presidential election and in the Brexit election. But I think the claim of sort of limitation of change or limitation of impact that I made a moment ago about algorithmic bias doesn't apply in that case. The possibility for continued expansion of data, surveillance, and better algorithmic tools to more and more precisely manipulate human decision making in domains where we still operate on kind of ideological presumptions of autonomy, as foundations of democratic decision making, as foundations of actually valorizing market-based decisions and so on, I would say that continued development of those capabilities poses challenges to human organization and institutions that become transformative and novel in character by virtue of continued incremental development of the capabilities. And it's getting some degree of disciplined handle on the nature of those transformative impacts that really defines the core of my interest in AI PULSE, and I think of those of my colleagues in the project as well. And I would say those are different both from superintelligent AI enslaving or exterminating human beings and from these credit-granting decisions are both opaque and thus violate due process concerns and racially discriminatory in their outcomes as they're already being applied right now. And so we need to find GDPR2 or, California Consumer Data Act 2 to address those.

I think that the power of artificial intelligence in bias is not one dimensional in that you could put it in between two levels of biases currently seen in the world without AI, but it amplifies what's currently going on. It makes it possible to make mistakes and do the wrong thing faster and at scale. And it's not necessarily amplifying human bias, it can be accidental as a result of missing data. For instance, the Google Photos tagging, where it tagged a picture of black people as gorillas wasn't racial bias at Google that caused that, it was the absence of sufficient pictures of black people.

Photographs of dark-skinned human beings in their database. Yep.

Yes, which could have a human analog, but might perhaps be a unique example of bias as an emergent phenomenon from big data?

I don't know if it's unique. I mean, it might be intrinsic to the search for new areas of application. Because if you assume that you're working with systems that have to be trained upon large volumes of historically available, tagged data - I mean, I know there are advances underway in sort of learning algorithms that require less direction based on tagged data, but most still do - that puts a kind of a historical conservatism into anything that decision algorithms are likely to do. And I think it actually creates all kinds of landmines for serious errors as the system gets applied in domains that are different from the domains upon which its training data came from. And in specific identified instances, such as that Google Image tagging failure, once the omission and its resultant problems are identified, then it's obvious what to do about it. And it's not that it is that difficult to generate large numbers of images of more diverse collections of people - all colors and ethnicities, and dresses, and customs and so on. The problem is that to the extent that general capability is expected or imputed to the systems, it might be a systematic pattern that they are repeatedly moving into areas for which their training data are not sufficiently relevant and there is an increasing risk of introducing certain kinds of high stakes errors.

Right. And I don't have a current update on this, but at least three years after that error, Google still had not addressed this issue beyond removing gorillas as a tag from their algorithm. And they were soliciting pictures of ethnic people on the streets, according to some accounts, which surprised me that the data they want shouldn't have been as available in the internet already. But that's not something that I've got some really current information on.

I don't know the current status of that. And of course, that particular example was egregious in a way that provoked a huge, hostile response based upon the disrespect and the insult. I mean, it wasn't that there were severe material consequences that fell to the person whose spouse or child I think was incorrectly tagged, so I basically don't know what's happened subsequently. It may be that their subsequent search for a more diverse set of human images is just further disclosing all the previously unknown biases in the data sources that were already available. And of course, accumulating primary data is a big job. It might be one aspect of developing AI-based systems that is not subject to huge economies of scale. And thus, I mean, there's this commonplace trope in AI debates now when people worry about concentration of power and antitrust and monopolization and leveraging of existing power to broader mechanisms of power, that the constrained resource for enterprise isn't capital anymore, it's data. And data is going to be subject to much stronger concentration and economies of scale and scope than capital ever was. And so the shift to the important resource being data is going to be a shift toward more and more natural concentration of economic power in the economy.

And we've been talking here pretty much exclusively about social rights and legal responsibilities, but the charter of PULSE's connections between technology, science and law, and with AI that can extend to its use in law offices. And there are cases where AIs are better than humans, and certainly faster than humans in vetting non-disclosure agreements, for

instance. So there's a huge amount of territory there. Is that within your purview as well, or more--?

It's not firmly out of scope, it hasn't been a priority area thus far. You can conceptualize the relationships between AI and law in a few different ways. And so one is AI-based systems transforming the practice of law and the education of lawyers - how do law firms work? What their cost structure? What does it take to deliver high-quality legal services? And how do you train students to deliver those? That isn't quite identical to broader concerns about the impact of AI on labor markets, but it overlaps with them to a substantial degree. To a substantial degree, what's going on there is automation reducing the need for human judgment and decision making in various functions that contribute to the economy. And so the lawyers are in the same position as the stockbrokers were and as the pharmacists are going to be and as the money managers and investment advisors are going to be. There's this irony in automation that AI-based and data-based automation that targets thinking jobs, rather than physical skill jobs. And so ironically, my grandmother's advice that a skilled trade was always a reliable way to make your way through life is oddly becoming true again after some decades of "No, no, go to law school." So there's the labor displacement aspect, and I guess I want to submit that that actually is not that distinct in law. So if you're running a law firm, you of course want to be up on all the current generation of tools that are offering to improve the output of your office, reduce your costs, etc. I think other aspects of the relationship between AI and law actually pose deeper challenges and are more interesting and more distinct. So the automation of authoritative decision making. Authoritative decision making is conducted by not US states, but small state organizations. This is huge in its implications and we see it kind of creeping in at various margins. So, use of machine learning based systems to adjudicate immigration, citizenship and refugee claims, for example. There's been substantial deployment of those systems already in the Government of Canada, for example, in employment and citizenship Canada. We talked about criminal justice already. So AI systems that, although formerly in advisory roles, given workload constraints, end up exercising very substantial decision-making authority over decisions of huge stakes to human beings - criminal justice, health care, education. But then also, people conceptualize these issues mostly in terms of challenge and risk, but there's also a kind of a positive aspect and opportunity aspect to them. I don't know. Consider sort of interminable disingenuous fights over legislation based upon contending incompatible claims about the budgetary impacts of this or that draft legislation. Imagine if the central process of impartial adjudication of impacts of specific statutory language have some element of expert-based automation to it. So, could AI and machine learning systems take over some of the task of the Congressional Budget Office or the GAO in terms of budgetary scoring of legislation? Would that be okay? I mean, would that breach sort of constitutional or procedural protections that are important to keep on those processes or would it be an improvement? Could AI systems somehow manage mechanisms of citizen and stakeholder consultation to help develop collective decisions around regulation, legislation, or program design in a way that would be more reliably reflective of popular sentiment than current incomplete and impartial mechanisms are? You can get as radical as you want. So I mean, I tried to sort of throw a long Hail Mary pass forward with my piece on AI central planning of the economy, but my colleague and co-founder of AI PULSE Richard Re, is working on a project

called AI Overlords that basically poses the question, “How extensively could AI-based systems fully supplant state decision making? And in what respects is that a desperately terrible, frightful thing and in what respects might it to be a constructive thing?”

Well, we’re talking around a number of impactful subjects here. I think if you title something “AI Overlords”, you’ve kind of forfeited your right to not be sensationalized in the media.

Yeah, but it is only his working title. And I think his aim is that there’s a little bit of irony in the title, because of course, yes, it is sensational. And it is also strongly pejorative, right?

Right. Some people can’t take a joke.

Sadly, no.

And when you ask a question, “Could AI run the economy better than the markets?”, I feel I have permission to ask, could an AI run the courts better than juries?

Or better than judges. Actually, my colleague, Eugene Volek, who loves to stir up debates with provocative contributions, he has a paper a couple of years ago called “Chief Justice Robot”, which basically poses the question, “What do we expect of judges?” On what criteria of performance do we evaluate judges to say this person is a good judge and this one is a less good judge? And if we imagine that that specific decision-making function was replaced with an automated algorithm, rather than a human being, would that be intrinsically problematic? Or can we imagine that it might not both, one, render better decisions according to consistent criteria, and two, on that basis, be permissible even if we start with sort of prior objections, saying, “No, no, decisions of this gravity must be undertaken by only human beings”?

Well, I can think of numerous decisions by white judges and juries in the South that could have been improved upon by AIs.

Right. I mean, when we started a few minutes ago, talking about use of machine learning based systems in criminal justice, generally speaking, the issue is that they are biased, that they exhibit racial class and other biases, but they do so somewhat less than the alternative pre-existing human decision systems. And actually, one area in which I think there’s some real progress is that people doing the training algorithms for them, I think, are getting skilled at reducing the severity of those biases by tweaking the learning algorithms and the training data sets. Of course, nobody exactly knows what a completely impartial, unbiased criminal justice sentencing decision system would look like and so nobody knows how to evaluate how close to where that aim a given system has approached. And of course, there is a morally consistent position, which is that any such bias is impermissible and any system that exhibits such bias must not be used. And don’t bother me with questions [of] compared to what!

That’s the end of part 1. I cut the interview off there because Ted then branched off into a parallel with autonomous vehicles which led to a discussion that took a large chunk of time that I want to keep together.

I find it perpetually fascinating how people working in this field find themselves ping-ponging between immediate consequences of AI, like bias, fraud, privacy, and explainability, and long-term consequences, like job automation and controlling machines that accumulate increasing levels of decision-making power. You heard Ted there start off talking about how early discussions on the impacts of AI were dominated by grand themes of superintelligence and the Singularity, and he helped bring it back to the here and now with what's going on right now that needs to be addressed. Yet PULSE also addresses long-term concerns, including the possibilities of automating the legal system and redesigning the economy. We rapidly bump up against our painfully inadequate powers of prediction and how little we know about the forces that will shape the future. So my hat's off to Ted and AI-PULSE for tackling that, and also to the Open Philanthropy Project for their funding of PULSE.

In today's headlines from AI, the big one is Google DeepMind's AlphaFold breakthrough in protein folding. Molecular biologists are not given to florid overstatement, so their reactions to this make it all the more pivotal. This is on a par with DeepMind's AlphaGo, which in the space of a couple of months advanced the level of computer ability in playing the game of Go by a degree that was expected take at least ten years.

Okay, but first, what is protein folding, why is it important, and why is it so hard? Protein is not just something you get from dairy. There are all kinds of proteins – over 200 million at last count - and they are big molecules made up of chains of amino acids. Proteins make up bigger structures, and play important parts in the functioning of our cells. They also appear in viruses; for instance, the coronavirus has spikes that are proteins. We can analyze a protein to find out its sequence, which is to say, the order of amino acids in it. But the other thing that's important about a protein is how those chains are folded up into a 3-D structure, because the resulting shape determines how it behaves. Like, for instance, whether it can grab onto the spikes sticking out of the coronavirus and block it.

We get only indirect information about that structure. To figure out what its shape really is we need techniques like cryo-electron microscopy, nuclear magnetic resonance, and x-ray crystallography, which can take hundreds of thousands of dollars and years of trial and error for a single protein. Some proteins remain unsolved after much longer. The number of ways those chains can be folded is so huge that trial and error is nearly hopeless.

But now DeepMind's AlphaFold, using 128 of the latest generation of Google's tensor processing units, can predict the folding of a protein just from the sequence, and it is unbelievably good and fast at it. I say unbelievably, because researchers who looked at its results thought they were so good that maybe it cheated. So they gave it a special challenge: To find out the structure of a membrane protein from an ancient species of microbes called *archaea*, which had so far defied our attempts to model it. AlphaFold returned a detailed image of a three-part protein with two long helical arms in the middle. "It's almost perfect," Dr. Andrei Lupas, Director of the Max Planck Institute for Developmental Biology, told *Science* magazine. "They could not possibly have cheated on this. I don't know how they do it." Mohammed AlQuraishi, a systems biologist at Columbia University said, "It's something I simply didn't expect to happen nearly this rapidly. It's shocking." He thought it would take 10 years to get from the 2018 version of AlphaFold to the level of results they have now. Dame Janet Thornton at the European Bioinformatics Institute in Cambridge, UK, has been working on proteins for 50 years. "That's really as long as this problem has been around," she said in a press conference. "I was beginning to think it would not get solved in my lifetime. AlphaFold will open up a new area of research."

This is huge. I've been frustrated by much of AI research for a long time because all it seemed to be used for was learning consumer preferences to make recommendations on what they should buy next. I've felt like paraphrasing Peter Thiel: "I was promised the end of disease and pollution. I got a suggestion to buy barbecue tongs."

But AlphaFold now starts to deliver. This level of breakthrough in protein folding really can make substantial advances in curing diseases, in creating enzymes to break down plastic waste, and devising means to capture atmospheric carbon to reduce greenhouse gasses. Way to go, DeepMind.

Now I want to put in a plug for my upcoming continuing studies course that covers the same theme as this podcast: What is AI, Why will it affect you, and How can you and your business survive and thrive through the AI revolution. And this course will again be online! For the obvious reasons. So it will start at 5:30 pm Pacific time which means it will be at a reasonable hour for everyone from China, Australia, Japan, and Hawaii through to New York. Also including you folk in India – I see you on my stats page. I said last time I would make it easier for you on the next course and here we are. Sorry to those of you in Europe and Africa this time.

This is 10 *hours* of instruction. It takes place over 5 classes, one per week, starting February 10 and running through March 10. There's a registration link in the show notes and the transcript. Or you can go to continuingstudies.uvic.ca and search on artificial intelligence, it's the first hit. You'll know you're in the right place when you see University of Victoria at the top. Since it's online I'm hoping they won't cap the signups at the limit they have on the page. I get paid the same no matter how many people are in the course, the reason I want lots of people there is to get more of my message out there and because the more people the more fun and impactful the classes.

What are we going to talk about? A huge variety of things, from the history of AI, to the present issues, to the speculative future; from the people who are influential in the field to the impact of AI on jobs, media, and society. We'll spend a great deal of time explaining AI at a practical level – that doesn't require computer experience – so you get a good idea of just what it can and cannot do now and in the future.

Obviously that's a really broad syllabus. Just like this podcast. We're not going to teach how to program AI, there's no code or math, it's all about... well, AI and You. *Everything* I'm producing is doing that job: this class, my videos, my TEDx talks, my book, and this podcast. Next up will be the 21-day diet plan.

My vision – just so you know where I want to take this – is to produce not just a course but an entire department, giving multiple programs of multiple courses, for credit, at undergraduate and graduate levels, and also with high school and corporate training versions. The idea is not world domination – well, not just world domination - but to create part of what we need to help people understand how to deal with, and leverage, disruption.

<https://continuingstudies.uvic.ca/humanities-and-social-sciences/courses/artificial-intelligence-and-you>

In next week's episode, we'll conclude the interview with Ted Parson, starting with a fascinating discussion about regulation and autonomous vehicles, including branching off the infamous Trolley Problem to ethical and liability implications, and some of the PULSE project's work on imagining the broader long-term consequences of artificial intelligence development. And if you're embarking on a new career or career change, or you know someone who is, we'll discuss what working on the project would be like specifically with you in mind. 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>