AI and You

Transcript

Guest: John Zerilli

Episode 78

First Aired: Monday, December 13, 2021

It’s pouring with rain outside, some of it may get on the soundtrack; we don’t care, it’s episode 78! Today’s guest is John Zerilli, a philosopher with particular interests in cognitive science, artificial intelligence, and the law. He is currently a Leverhulme Fellow at the University of Oxford, a Research Associate in the Oxford Institute for Ethics in AI, and an Associate Fellow in the Centre for the Future of Intelligence at the University of Cambridge. So, straddling both sides of the Oxbridge divide there, that’s a bit of a balancing act.

He has written several books, including *The Adaptable Mind*, in 2020, and *A Citizen's Guide to Artificial Intelligence*, which just came out a few months ago. With a title like that, you can imagine that it’s intended for people just like you, the listener of this podcast.

It starts out with a chapter describing what AI is, then each chapter, some of which are written by coauthors, tackles a major issue or theme in AI today: Transparency, Bias, Responsibility and Liability, Control, Privacy, Autonomy, Algorithms in Government, Employment, Oversight and Regulation. It reads like a manifesto or a call to action for the layperson to get informed.

So we’re going to start out talking about what sort of results or influence he was intending to get out of publishing this book, and how some of those issues impact the average person today, in particular, privacy in this first half. Let’s get to the interview.

John, welcome to the show.

Thank you for having me.

We’re here to talk principally about your latest book that’s come out in 2021, *A Citizen’s Guide to Artificial Intelligence*. It’s a really interesting title because I know from reading this that you are very deliberate in your choice of words. So, speak to the *citizen* aspect of this please.

Because the book’s cover obviously has seven authors on it, it immediately inspires the thought, “Oh, this has been written by a committee. There are too many people with too many different and potentially jarring styles, so this will not be a friendly read.” To counteract any suspicions that the authorship might have aroused, I thought, we need to have a title that conveys the true spirit of what this book is about. Hence, I used the word “a guide” and I also used the word “a citizen’s guide” to say it is for anybody who is interested in the idea of artificial intelligence and what it means for us. So it’s basically addressed to the interested lay person. The word ‘citizen’ also gets at another dimension of the audience, which is, I want to address the lay interested people in their capacity as a voting citizen of a presumably democratic polity. So, I want to highlight the fact that this book’s sweep will be necessarily political.

You definitely talk about it being political. You also emphasize the need for vigilance near your conclusion, and when you're talking about vigilance and you've already set up for the issues, the reasons why vigilance would be needed, but who are you addressing there? Is it more the lay person or would it be policy makers?

It still is the lay person, but in a way it's both, because some lay people will also be public servants. I mean they are lay people insofar as they don't know much about artificial intelligence; they might have been students of public policy or economics but they may not have much knowledge about machine learning, artificial intelligence. So we can still call them lay people, but I'm addressing I suppose lay people in the political process more actively and those that are not so actively involved. Because the idea is to get anybody, insofar as they are a citizen, to think about what was being presented in the book and think about how they might pull the levers that they have available to them to get the system working the way it should be. Whether that's by contacting a member of parliament, agitating for parliamentary inquiries, bringing things to the attention of those that are in power so that a momentum can get underway. For change where it needs to happen We're already seeing that all around thw world.

I described for the audience, in the prelude to this, the sections of the book about the issues that many which we're familiar with on this podcast by now about bias, responsibility, transparency, algorithms, use and employment and regulation, and when you talk about the getting a handle on these, asking people to be activist in some form here, to claim a role. I know this is a very hard question, because I've been trying to answer it for years and asking a lot of other people. Where do you see those levers that they can pull? And there can be different answers for different groups of people. But I know for instance in the UK you've got the All-Party Parliamentary Group on AI which is been looking at this for years in broad strokes and they're certainly very well-informed at this point. I think they're all struggling with this idea row of "who should be involved and how should they do that, and what's the role of regulation?" And regulation isn't the only government tool, certainly there are incentives; you can push in both directions. Aside from government, give some examples of how people might take to heart your message and change what they're doing.

You're right in pointing out that different people have different circumstances and will be able to engage in the political process in different ways. So the general message is just whatever ways you are able to engage with, do grapple with these issues. An example for someone that may not have much to do with big tech, machine learning; may not even be involved in government, might just be the simple act of refusing to allow cookies when they visit websites. It is an extra step; but since I've implemented this change myself in my own browsing life it actually doesn't add much more than about three seconds. Because now because of the GDPR, the main legislation governing data protection in Europe and countries that transact with Europe, there is a requirement for these websites now to be very upfront about cookies that are being put on your device, and it is actually being streamlined to a surprising extent. So now when I visit a page it'll say "this site uses cookies you can either agree, accept," or it'll have something like manage cookies or words to that effect. All it literally takes for the most part, I would say 9 times out of

10 is just to click Manage cookies and then Reject all. Basically. And then you can go on doing what you do. Even that simple change means that you are contributing in a smaller way to creating an ecosystem where Big Tech doesn't just assume that your data is there for the taking. And I know it sounds miniscule, you know, "one person's contribution." But it's not unlike the case of, if you remember back to the late 1980s and early 1990s, I was just in school – I won't tell you quite how little I was. I even remember campaigns to encourage people to dispose of their rubbish when outside. At one point you had to tell people not to just throw litter wherever they were, if they went into a park and had a picnic. You had to tell them to take their wrappers and plastic and dispose of them. Now we kind of just take that for granted. It's a small change but it's one that it has made a big difference to the appearance of many of our urban spaces. For those that don't have that contact with machine learning and deep data and deep tech there are simple steps like that I think will go a long way towards changing the information that we give, because at the end of the day, the entire phenomenon that we're experiencing, this renaissance of AI in the form of machine learning runs on data. So, any dent at all in that is something. I don't know if it'll be enough, pretty sure it won't be enough to deal with the problem but it's something that is empowering.

Is it specifically privacy here that we're discussing, or does it leak into other areas?

I'm thinking mostly in terms of privacy, because the information that you give away online - the problem with it isn't just that you're giving away your information and that might be somehow weaponized against you personally. The problem is that by giving information away you are giving a very powerful resource to companies that use that information to run their platforms, and these platforms can be put to nefarious ends and they have. The entire business model on which Facebook runs and which Google runs requires extracting as much information from individuals as possible. And what that results in, I mean, just to take the case of Google and Facebook, is it feeds a system that has effectively become the public square. We're no longer in Renaissance Italy where we can go out into the piazza and debate political issues of the day with all and sundry. We basically, our contact and our exposure to competing views are now almost entirely mediated with online platforms. And so what you see on Google is a function of what Google essentially decides to show you when you enter a search term. And the news that you see on Facebook in your newsfeed is a function of what Facebook's algorithms have decided that you really ought to see. So, the material that you see in that public platform, the modern-day equivalent of the old piazza, is almost entirely governed by these two corporations. So by you not giving them that information, you are in however small a way, restricting their power. You are saying No to that model whereby they get to determine what the public sphere consists in.

That's a very good point. You give us a suggestion, among many others, about how to find in Facebook where it has decided what categories of advertising it will show you, and it actually lists those, which surprised me, because I thought it would be dynamically-generated machine learning list where they wouldn't be able to characterize the clusters that way. But no, they show you these specific categories of advertising, and there were probably about 50 for me, some of them predictable - artificial intelligence - some of them not so, like I guess that came off from some search or something I watched once but it really has nothing to do with me. And you can manage them – there also seems to be this principle on Facebook that the more personal the information, or the closer it gets to your privacy, the more clicks it takes to find it. And this is very timely here when we've got the Facebook Papers being examined and talked about in the US Congress at the moment. I want to flip this at the moment, because that's a narrative that we've heard a lot about, to ask, in order for AI to serve us, to do what we want, it's going to need information about us. It's going to need personal information about us to do personal things for us in the same sense that if we want a digital butler it's got to know our preferences the same way a real one would knows how many sugars you take in your tea. What sort of governance do you think could handle that responsibility, where you would want to take your foot off the brake?

Could you clarify that question? Are you asking what sort of governance structure would best protect individuals from having their personal information used to personalize content?

How could AI be developed and deployed for our personal use where we could trust it with that level of personal information to realize the benefits of it having that personal information?

I think probably it's going to come down to the privacy regimes to which we're all subject. Basic principle of privacy law is that information - you can consent to the information being used but it has to be consent to named and specific purposes. So, if you look at the terms and conditions of, say, how Facebook gets your consent. You're basically consenting to them using that information, as they put it, to improve Facebook o to improve our services. That's not a named and specific purpose. That's using your information to implement Facebook which is almost like a tautology. Clear up more precise and firmed-up privacy laws will see to it that you cannot give consent to such ambiguous causes. If Facebook wants to use your information to do something very specific, so, for example, to figure out the kind of person that enjoys long train journeys, then they have to tell you that that's what that information will be used for. It will be used to train the system to determine who likes long train engines. And maybe they'll find a certain type of a demographic, say white, maybe Christian, over the age of 55, might like to take long scenic train journeys. Well then, we need to be told that, and that's what we'll be consenting to. If you just consent in the abstract to something vague, it's really no consent at all because consent is nothing unless it's knowledgeable consent. But at the time that you give consent if it's for such a vague purpose as to improve our services or improve the quality of our platform, you don't know what that means. That could mean all sorts of things. In what way does that mean improving the platform? Does that mean ever more invasive insights into the kind of person you are? To the extent that they might be able to tug at you with ads that may press your buttons in some way as to push you one way or another in an election. That might be improving the website as far as they're concerned, but you don't know what you're consenting to.

And those are the kind of penalties that we've paid for this. It's become a widely dispensed truism that if you're not the customer, you're the product, and we're not paying for these services, therefore we are being used. But I don't want to let go of the idea that someone somewhere could make this digital butler available to me to be a personal assistant, because I would very much like one, but I can't afford a real one. And so it would have to know a great deal about me that I wouldn’t want abused, just like any personal assistant would. I’m prepared to pay for that and in order to do that we have to have the assurance of some appropriate degree of privacy. But now I wonder if the presence of these huge oligopolies, Facebook, Google, Amazon, and Apple giving away AI for zero privacy; does that crowd out the possibility of the kind of service I would pay for that would have the guarantees of privacy?

I think it might. I think it might. I mean the digital butler as you put it may just not be worth it. Because I don’t know if there is a way to protect personal data to the point where you can give a company so much intimate knowledge about you without running the risk that it will be sold to someone else or in some other way find its way elsewhere in the data economy. Even the NHS, without getting patient approval in the UK, was willing to give Google patient data and of course it might be sold in terms of the benefits of being able to mine so much health data would confer to future generations in terms of medical insights and discoveries. But that’s a bug in that we should be involved, and the British public wasn’t allowed to come on board. They didn’t give consent to that handover of information to Google. So, I don’t know if the digital butler is actually possible. If no less an entity like the NHS, which you would think would really have extremely high regard for our private information, was willing to give away that information to a major player like Google. I think the butler might not be feasible.

I think it might lie on the other side there might be some sort of breakthrough in data compartmentalization and security and was that NHS data anonymized before they turned it over?

It probably was. I think it was. I’m not sure I’m familiar with de-anonymization techniques but it’s not too difficult.

Yes.

It’s not too difficult if you really want to find out whose information something belongs to, whose information it is it’s easy enough to do.

I think I saw that combining something like four or five demographic data types would be enough to deanonymize data to recover identity. You bring up issues in the book of bias with some specific examples about AIs making decisions based on characteristics that they’d harvested from data that weren’t in the traditional classes of protected attributes of protected classes but nevertheless make us think about this. Like could AI that’s in charge of deciding a court case decides that part of its decision is based on what someone has for breakfast which if it was exposed in a human Court would be tossed out as irrelevant. But that is exactly the sort of thing that an AI might come up with in its cluster analysis if we unpacked it to see what was there. If you look by analogy at the way that AI is trained on image recognition, and say sheep, pictures of sheep, this has been demonstrated by people like Janelle Shane relentlessly and others, that they are recognizing them not just by the boundaries of the white fluffy things but by the fact that are in landscapes. So a landscape that looks like a sheep field will have a significant chance of being tagged as containing sheep by an AI even if there’s nothing in it and so some sort of cluster analysis of factors used by an AI in a court case might well have something in there like breakfast food and your other example was, we would we be justified in withholding a loan to someone because they like fennel - I don't like fennel; maybe that's a plus or a minus. Fennel likers you say aren't a protected class, but it still looks like discrimination, so we wouldn't want that to happen. But what if in these cases somehow it turned out AI was right, or we couldn't prove that it was wrong? That we just hadn't thought about what the connection might be between fennel and loan qualifications. Are we looking at this the right way?

Yeah so, I leave that as an open question at the end of the prologue. It's precisely the kind of question that the rise of machine learning algorithms in all sorts of areas, of public and private decision-making poses. And I still don't have an answer to that. But I think what we can't do is to necessarily dismiss the result of a machine learning algorithm just because it seems unintuitive to us. It's a point that's being made by the scholars Selbst and Barocas in the US and they talk about the fact that just because something is unintuitive to us doesn't mean it needs to be dismissed out of hand. This is unlike the case with a human. If a human were to decide something on the basis of such a spurious correlation like fennel and credit-worthiness you would have no hesitation in dismissing the decision as being unjust, not even a matter of discrimination, it would just be an absurd irrational result. But that's because we know what humans are like and we know roughly how human tick, whereas with the machine that comes upon the same correlation we can't be so sure that it's just irrational.

It could be like AlphaGo's *Move 37* which no human Go player would have played and threw everyone especially Lee Sedol for a loop but later they concluded that was brilliant.

Exactly. So, what we have at the moment are legal regimes which for the most part exclude such results. So, a result like that that associated some random feature of the person with creditworthiness would just be under current laws of what it means to be reasonable, a reasonable decision-maker would be considered illegitimate. But the point that I put in the book at the end of that prologue is perhaps the default position of outright dismissal and rejection needs to be revisited in the case of machine learning. So I don't have an answer yet as to what we do about that but that's what I'm currently engaged in thinking about in my current project at Oxford.

That's the end of part 1 of the interview. I was particularly taken by the idea that there could be something that looks like discrimination – against fennel lovers in this case – that isn't legally discrimination, doesn't really have a moral basis for arguing that it's discrimination, but it still doesn't feel right. Even though it could be that there's some connection between liking fennel and the outcome the engine is being trained for, and we were too dumb or too blind to see it, but the AI found it. Even if it's not fennel, we should expect situations like that to crop up more and more.

In today's news ripped from the headlines about AI, Alexa has been outed. The identity of the original voice behind Alexa has been revealed to be voice actress Nina Rolle. According to journalist Brad Stone in his new book, *Amazon Unbound: Jeff Bezos and the Invention of a Global Empire*. Neither Amazon nor Rolle would confirm this, but you can go to Nina's website and hear her for yourself. Of course, she wouldn't have recorded every word that Alexa can say (I had to unplug mine for this recording otherwise she'd be all excited about hearing her name and interrupting – What is it, Peter? I'm sorry, I don't know that one), but enough phonemes for machine learning to learn how to construct them for various texts' inputs. What I don't know is whether they engaged her for the seven other languages that Alexa speaks. To me they sound like they could be the same or a different speaker.

Next week we'll conclude the interview with John Zerilli, when we'll range over ethics, education, bias, and cognitive science. How does looking through the lens of philosophy – and we've had several philosophers on the show now – inform a deeper perspective of AI? 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