accenture

AI LEADERS PODCAST RESPONSIBLE AI

AUDIO TRANSCRIPT

Dr. Adrian Weller [00:00:00] It's better for society if we can form good, comprehensive governance so that all companies have the right kinds of incentives to do the right thing.

Ray Eitel-Porter [00:00:15] Hi, I'm Ray Eitel-Porter, a managing director in Accenture's Applied Intelligence Practice and our global lead for responsible A.I. I'm here today with Dr Adrian Weller, program director for A.I. at The Alan Turing Institute where he's also a Turing Fellow leading work on safe and ethical A.I. Really delighted to be speaking to you again today, Adrian. I know you have a particularly interesting background combining your experience in industry and then moving into academia. What was it that got you interested specifically in the ethics of data and A.I.?

Dr. Adrian Weller [00:00:54] Gosh that is a great, a great question. First, Ray, it's great pleasure to be with you today, and I'm really looking forward to discussing these topics with you. Your expertise and experience are really valuable in this area. I think I've long been interested in issues around ethics and justice and fairness, those sorts of topics, but I can remember there was one particular event that really made me think that perhaps it was a key time to start thinking about these issues. That was at a small conference that was organized by the Future of Life Institute in 2015, I think, in Puerto Rico. It was a great event.

There were only about 60 or so people, but some really great high-powered people, and it felt like a bit of a turning point in the way that the community was thinking about the risks of artificial intelligence both in the near-term and in the long run. So, there were people emphasizing the long-term dangers, people like Nick Bostrom saying that we need to be worried about those kinds of issues. And there is, of course, quite a wide range of opinions about those sorts of points. But in addition, there was some really interesting discussion about issues which were coming on the horizon at what felt like quite a clip at the time. So, for example, it was a really interesting discussion around autonomous vehicles, and I remember at that time - remember it was 2015 - I don't think at that time had had any actual fatalities, happily, at that moment from autonomous vehicles or cars. And the discussion then was 'well in the long run or even in the medium term, this might be such a valuable thing for society that should we think about whether we should incentivize companies to take the risk to develop these kinds of vehicles technologies, recognizing that when someone has a casualty that could really push the whole industry back a really long way because it will be such a public outcry that the whole thing might be shut down for a bit and there might be tremendous liability issues so should we perhaps limit liability for those companies in a way that I believe is similar to what was done with nuclear power

stations early on in their development. So, everyone thought nuclear power could be very good for society but what if there was a big nuclear accident that could be really bad, so you had to protect companies so that they felt able to develop that technology. Should we do something similar with autonomous vehicles? And I think it's been really fascinating in the years since then that we have seen some fatalities in autonomous vehicles, but public perception has shifted so much, arguably so quickly in the few years, that it hasn't really stopped the development of that technology at all. So, I think it's a good lesson to observe that public perceptions and attitudes can change quite quickly.

Ray Eitel-Porter [00:04:05] That's interesting, and I like the point you make about your own passion for ethics and justice. I think that is, to me, one of the facets that I really like about this particular sector and topic that we're working in. Everybody is highly engaged, you know. Everyone who is working in the space is doing so because they have a deep-seated belief that this is the right thing to do and that it's a very important aspect of artificial intelligence that we need collectively to be working on. And I guess given that, why are we still seeing challenges for organizations in formulating good governance and putting things into practice?

Dr. Adrian Weller [00:04:48] I think there are lots of challenges, and I'm sure we'll get into quite a few of them. One of them is what should good governance be? So, before you can even put it into practice you need to know what is it. What should it be? I think that's quite tricky. Companies clearly have various pressures on them to perform. They have a quite understandable pressure from their shareholders and from their management to try to make profits and to be successful. And I think overall our market economy has done a very good job at helping technologies come to the fore that have been useful and beneficial for society, but we should know that companies have these incentives to develop products that will make money and, in many ways, that can result in things which are good for people, but sometimes it can lead companies to do things which are not necessarily in the best interests of society or even in their own customers in the long run. Perhaps one good example might be social media platforms, where clearly, they have an incentive to try to attract people's attention, and it's difficult to measure people's attention very… you know...what's difficult to really know what sort of thing you want to measure. An easy thing to measure is if people click on something and a whole advertising business is growing up around that. So, it's quite natural to start thinking about optimizing systems in order for people to want to click on them. And if you give people things that they'll click on, you could argue that you're giving them what they want but I would say that it's perhaps a bit shortsighted or, putting it another way, it's giving people what their very short-term instincts might make them want to click on, a bit like if you give them a sugary sweet to someone. They do want it, they do like it, but it might not be in their best interests, and it might not be in the best interests of society. So, I think when you take a step back and try and think carefully what's going on with companies, are they, in a sense, producing externalities on society which we're not capturing or governing effectively? How can we think together as a society about what sort of things we want, what kinds of constraints might we want to place in order that systems are developed in a manner which is safe and ethical? But there's a lot of discussion around that, as you know, and as you're involved in.

One of the things I'd highlight is that we've seen many institutions come forward with sets of high-level ethical principles. We've seen well over 100 now, and the good news is that there's quite a lot of agreement about the sorts of things that those ethical principles highlight, things like fairness, justice, transparency. But I'd suggest that it's actually very difficult to know exactly what those terms mean, unless you start looking at things in specific context and try and understand, well, what exactly do we want to provide for whom, why, and start thinking very carefully about how that should work. Happy to get into detail on that if you'd like.

Ray Eitel-Porter [00:07:54] Well, there's a few points I'd like to pick up on there. So, I'm just thinking about where we start maybe on specifically your point about fairness and transparency and then what I would love to do is come back and talk about the question of the extent to which organizations can sort of self-regulate, if you like, versus whether we need some external legislation or guidance or whatever, because of the almost inherent contradiction that you're implying there between the motives that a company may have to operate under. But perhaps let's start with the question of fairness and transparency. And, as you say, on the surface, the words sound reasonably clear. But actually, as we know, there's a lot more going on beneath the surface when you try to actually put those into practice. Can you explain and bring that to life for our listeners, please?

Dr. Adrian Weller [00:08:47] Yes, perhaps let's talk first about fairness and then turn to transparency. I think fairness is a thing which all of us think is a great thing. We all want a fair system. We all want fairness. But what each of us thinks is fair can vary from what someone else thinks is fair. And so, when we're talking about the ways that companies should behave or that algorithmic system should operate, we're usually talking about how a system should operate and considering how it impacts different groups of people. So, as you vary your view of fairness, you are typically shifting the benefits and the cost and the risks across different groups of people and what exactly is fair for everyone is a hard thing to really figure out. And actually, it's often a political issue, political with a small p. Different people have different views, and that's totally understandable and totally reasonable, you might say totally fair. And one thing that I think is really important is that we have to hear those views. You have to make sure that we listen to people. So, the first thing to do, I think, is to engage with users, practitioners, stakeholders in a particular setting and just listen to people's hopes, fears, concerns, understand how they feel about the issue. Once we've done that, it doesn't mean it's going to be simple to figure out what is the one fair answer. In fact, there might be various points you could get to just depending on how the discussion goes. One example that I think might be useful to think about would be the tax system. We all think we should have a fair tax system, but if you ask different people what they think is fair, you'll get very different answers so it's OK to have different opinions. But to start with, we need to listen to those different views and then we need to try to form some kind of opinion about what makes sense. As you know, the way that the technical community has been thinking about fairness, which has taken off in the last few years and I think is a really good thing for the most part, has been a little bit myopic and I'll say that as someone in that community.

But, you know, I think this community does recognize the shortcomings that the technical folks, particularly technical machine learning people or people who are building programs have. What we like is very specific clean measures, which we can then try to optimize. Fairness is often much more a complex, sociocultural kind of topic, and it's very difficult to make it very precise. But because we do like to make things precise and if we can get something in a precise measure, then we can build systems subject to those constraints. There has been quite a bit of progress made where you can choose from a whole menu - I'm sure you know about this Ray - there are over 20 different measures of fairness that different people have proposed, and you can design a system that will try to do well subject to the constraint of performing well on one of those fairness measures. Then, of course, it leaves a big question about, well, which of these fairness measures is appropriate, if any. It's often maybe a much richer set of challenges that are at stake, but this just starts to highlight some of the problems. And just to make it even worse, if you say, well, maybe I'm not sure if I want fairness measure A or fairness measure B, can't we have both of them. If you try to optimize an A.I. system subject to both of those constraints, you're typically actually constraining it far too much and you won't get anything that's particularly useful in the real world. So, we need to think very carefully about what really do we want from a system that requires engaging proactively with users and those who are going to be impacted. That's just the starting point.

Ray Eitel-Porter [00:12:21] I think that's a very interesting illustration also of the need to have good governance procedures in place in an organization that stemmed from or embody, if you like, the core values of the organization. Because we've seen that a lot of the impetus for improved responsible A.I. within organizations is actually coming from the frontline data scientists. And my suspicion is that part of the driving factor for that is the fact that data scientists don't actually want to be held responsible for making all those decisions themselves. Those are big, complicated issues, and they would like to have clear guidance and a framework that the organization as a whole has blessed and said, this is how we think about these different topics, and this is what our organization stands for. Here are the guardrails within which you should be making those decisions rather than just leaving it to an individual data scientist to make up oh well, in this instance, I think, you know, this definition of fairness is appropriate. So, I think it's encouraging to see that push for guidance and understanding that is really coming from the whole organization. It's not just a technical issue, it's a much broader enterprise-wide issue.

Dr. Adrian Weller [00:13:45] Yeah, I wonder if I could make a few points on that. One point is that responsibility, an important word that you just used, and of course, an important word that people are thinking about a lot. We've got responsible A.I. as a big topic. People are talking about safe and ethical A.I., trustworthy A.I., responsible A.I. I think that for the most part, when people say trustworthy A.I., they mostly mean the same sorts of things I mean as safe and ethical A.I., which is broadly speaking how can we try to ensure that systems which are deployed are safe and ethical and will not hurt people and will be generally good for society. Responsible A.I. is I think, a little bit less clear. I think sometimes it's used to mean the same thing, but sometimes it used to mean a broader thing.

There are lots of really interesting questions. One thing I would highlight is that you can think about responsibility in a way looking backwards or looking forward. So, when you talked about developers wanting to be responsible or not responsible - I'm not quite sure which would one they'd be more interested in - but there's a question about being accountable, looking backwards for what has been done and, in a sense, are you liable for things which go wrong? That's one kind of responsibility, and there are lots of interesting questions there: who is responsible if something goes wrong? Is that the designer, the developer, the deployer, user? Lots of interesting questions there. And I'll just comment very briefly on that before we come back to it, perhaps later, which is that I don't think there's a simple answer, actually. I think that we should have a legal system which certainly pays attention to what makes sense as who is actually responsible sort of philosophically, but also looks through questions like, well, if we set up the rules in a particular way, we'd like it to be reasonably efficient to try and figure out who is responsible under the rules, and we'd like to set it up so that the economics of it makes sense so that if you set up rules this way, people are appropriately motivated and incentivized to do the right thing. I think all that already requires a range of different disciplines to come together. A lot of the questions we're talking about, I think, requires lots of disciplines to come together, and I think that's a good thing, that we all need to think together at all points. Even as I have talked about responsibility looking backwards, looking forward, it's arguably slightly different. It's sort of trying to think, well, how can we try to really make sure I feel responsible to make sure this is going to work in the future, that I need to take a lot of care. I need to try to have guarantees of its performance or I need to test it really carefully and to think really hard about how to do that. So, there are those two aspects to it and on your point about, well, individual developers maybe shouldn't be forced to have to think about these tricky issues. It should be an organizational issue; I would say I completely agree. But I also think that maybe you need to go even further because if each organization thinks about this differently to another organization, then it's unclear how this will develop for society, and it's also not clear if we'll have the right incentives for companies together to do the right thing. Then one company may take a more lax view to its ethics and arguably might be able to get a stronger market position because of that, and we don't want that to happen. So, to the extent we can, we want to try and form a level governance playing field that incorporates the kind of ethical principles which all of us think makes sense.

Ray Eitel-Porter [00:17:02] I completely agree. Adrian, I completely agree, and that brings us back again, actually to that topic about can a company, if you like, self-regulate and deal with these issues on its own? Or does it need to do so in a broader societal context and what kind of guidelines are needed? What is your thinking about, let's say, regulatory control versus perhaps principles and guidance and self-regulation or other approaches? We've seen some organizations setting up, for example, independent ethics boards, which have more or less autonomy over certain decision-making and aspects. So quite a range of different approaches and, to this, where do you think we should be heading and where do you think we perhaps are heading?

Dr. Adrian Weller [00:17:55] A lot of interesting issues there. I believe that it's better for society if we can form good, comprehensive governance so that all companies have the right kinds of incentives to do the right thing. I think that makes it - as we were saying before - it creates a more level playing field, prevents a company from doing better by skipping corners. It also could be easier for companies, and I think it's interesting that we've seen some developments in this regard in the last couple of years. We've seen some companies saying we want to be told what's the right way to behave, and I think they do that to some extent. If you're being cynical, you can say it's a bit of a PR move. But I think aside from those kinds of ideas, there is this issue which we touched on five or 10 minutes ago, which is that no matter what position you take on being fair, there will be likely be someone who thinks it's not fair, that you'll be treating some people better than others so some people will always be unhappy. If some independent body has sort of laid down the law on how you should behave, then you can say, look, I'm doing everything that I meant to do and that sort of simplifies things and in some ways can protect companies, if that makes sense.

Ray Eitel-Porter [00:19:18] Yes.

Dr. Adrian Weller [00:19:19] One other important aspect of this that I would love your view on Ray is that increasingly we're seeing through mechanisms like ethics boards, but also various guidelines that are coming out. Companies, I think, are appropriately being asked to think through in advance how their algorithms or machine learning systems might be causing harm and to address those concerns and to document how they're addressing those concerns. And for the most part, I think that's a really, really good thing. But I do have a bit of a worry that a company might feel, even if they felt pretty confident that they were doing something in the right way, one, they'd be worried that they might be wrong and someone might come and show them that they should have been doing something a little bit differently, something they haven't thought of. But two, in some sense, that's kind of highlighting to the world these challenges and putting it out there, which can present a kind of target for people to shoot at the company and say, well, it's terrible, you're doing these things and you need to stop. To some extent, it could make companies retreat and try to ignore these problems, which is not what we want at all. We need to try and create some sort of environment where companies feel safe in coming forward and trying to do the right thing, but I think this is challenging to do, and I'd love to hear your views on it.

Ray Eitel-Porter [00:20:47] I think you highlight a very interesting and real concern there, Adrian, because I've definitely heard organizations expressing nervousness about putting something in writing on the record or what have you, which might potentially also in a legal situation, expose them to liability if they have recognized an issue. I think that some kind of safe environment, be it sandboxing and then some safe space that people and companies could engage with a regulator or a third party or what have you to discuss this almost under sort of legal privilege as it were. I think would be very helpful because otherwise, I think you're right, there is a risk that actually it pushes this sort of under the carpet and stops people from being transparent about what they're doing.

I think it's also particularly valuable because I imagine you would agree and we don't have all the answers yet in terms of how to fix, for example, fairness or explainability or many of these topics that we deal with. And I think it's only by collaborating that we're going to come up with ever better solutions and technical solutions to these challenges, so we want to actually encourage collaboration. Academics and others who are doing research need real world examples.

They need data and real-world examples of things that are working and not working. Otherwise, they won't be able to help us to fix the problems.

Dr. Adrian Weller [00:22:22] Completely agree. And more than that, we need real world solutions. So, we certainly shouldn't lock ourselves away in an ivory tower. We need to try, like you say, to work together and to see what works in the real world. It's very difficult to try to figure out exactly what mechanisms are going to get things right. So, I think, to the extent we can, it's good to be able to see some experimentation and explore what works. I think that's true nationally but also to some extent internationally. While on the one hand, in some cases, for example financial transactions, it's very useful if you can have harmonization of rules to allow trade and flow of thoughts around. In other cases, it's actually quite helpful for different jurisdictions to adopt slightly different rules and principles which correspond with their own local cultural values, and then we can see what works better, and I think that's a useful thing to do.

Ray Eitel-Porter [00:23:20] Yes, absolutely. And I think we're moving into the space here and as we discuss around the challenge between innovation and controls and constraints and regulation, that's often one of the pushbacks that one hears - oh if you bring in too many controls, et cetera, then you're going to prevent innovation and prevent us moving ahead. And you made the point earlier about the importance of a generally accepted societal view of this. I know one of the concerns that some people raised is that different regions of the world may move at different speeds, and that may put particular regions at a competitive disadvantage versus other regions. What's your sense on the feasibility, if you like, of trying to achieve some sort of global...is standards too big a word? But, you know, some sort of global approach to this that would help to mitigate that.

Dr. Adrian Weller [00:24:24] I think we are seeing moves to establish standards that are internationally accepted. You know about the Lee project and the ISO is working on a set of standards. The draft legislation from the EU suggests a set of standards established by certain organizations in Europe who would be used to certify different products, and I think there's a lot of interesting things to say about that if we have time later on. But on the question of whether trying to be responsible will be bad for the economy, well, we can see what happens, but I would emphasize the following points, which I think we should passionately focus on. If we get the rules right, even in our own jurisdiction, it will be better for society. We will be incentivizing the right kind of economic growth. We mentioned before the idea of externalities. If you don't prevent companies from generating the sorts of things that are maybe bad for society, they will produce a lot of that, just like pollution, just like we're having meetings around climate change. We want to try to stop companies generating problems for individuals and society in the long run and instead generating really valuable economic growth.

We can do that by getting the rules right. In addition, I think we help to remove some barriers to innovation. In particular, consumers, I think, can reasonably feel quite wary about embracing new technology because they're not sure if it's going to be good for them. If they can feel comfortable about that, that removes one barrier there, which is very helpful for adoption. And also, I think increasingly - I'd be interested in your thoughts on this - I think there's regulatory uncertainty which can prevent companies from investing in technologies because they're not quite sure how it's going to be governed. I think by getting this right, we actually have many ways that we're encouraging the right kinds of economic growth.

Ray Eitel-Porter [00:26:36] That's great. I have heard that argument before and I very much hope that is indeed the case. Interesting analogy I heard from somebody who said, you know, if you're approaching the edge of a cliff, actually having a fence in place is quite helpful because if you have the fence, you know just how close you can go to the edge of the cliff. Whereas if there's no fence, you actually tend to stay a bit further away from the edge because you're not quite sure you know exactly how far you can go and remain safe. I thought that was a good way of capturing the potential advantage of regulation.

Dr. Adrian Weller [00:27:16] I think it's a great way to think about it if you think about having guardrails to help protect you from going too far, but then I think another reason that's a good analogy is that it's that intuitive. You want to make sure that they are in the right place. If you have them too close, you'll be overly constrained. If you have gone too far away, things can go wrong. So, we actually need a lot of thought and attention into getting those guardrails in the right place. And again, this brings back to the theme we touched on before is that I think to do that well requires a lot of different folks to come together and communicate effectively. We need technical folks, policy folks. We need people from industry, lawyers. We need the public. I would encourage anyone listening to this actually who's interested to get involved and we all need to work together.

Ray Eitel-Porter [00:27:57] Absolutely. Maybe, on that note, Adrian, as a final question, what would you say are the highest priorities for us in this field over the next year or two?

Dr. Adrian Weller [00:28:09] Gosh, that's a great question, and I'd love your views on that, but perhaps if I may, I'll just I'll raise a few points. One is I'll come back to transparency, which we touched on before, but we didn't get time to really explore. I'll make a few comments on that because I think that's another area that people often cite as something that of course we need. And I'm a big fan of transparency. I spent quite a bit of my time working on methods for transparency. But I think it's important to recognize that transparency, actually, it's not always even a good thing. First, you have to know what kind of transparency you want. If you're talking about explainability of an A.I. system, who needs that information for what purpose? And think carefully about that, because that will then speak to what kind of approach might make sense, what kind of information you're trying to give them. If you're trying to help a developer of a system understand generally how it works, where it might go wrong, how they might improve it, that's one sort of system.

If you're trying to help a particular user understand why they were turned down for a loan, that's a different kind of system. If you want to regulate it to understand why did an autonomous vehicle crash so that you can help to prevent crashes in the future, that's a different kind of system. There are lots of different kinds of explanation and transparency and we need to think very carefully about in a particular setting what information is needed by who for what purpose, and then we can try to optimize that or try to work towards that. But also, I mentioned that sometimes it can cause harm. By providing a lot of information about the system, there are concerns about how that might reveal private IP about your system. It might reveal private information about the data the system was trained on, and actually if people are interested, there are more examples. I've tried to put all these concerns together in an article, which you can find on my website if people are interested. So, transparency is often really great, but it's not always the best thing. And also, it's not always the end the people want. Often, it's kind of a means to an end. People talk about wanting autonomous vehicles, where you can understand exactly what they're doing but which would you rather have? Would you rather have vehicles which are killing 100,000 people on the roads and you can understand what they're doing? Or would you rather have ones which are killing 1000 people on the roads and you're not quite sure what they're doing? It raises important questions around how should we try and test and ensure that these systems are safe or doing whatever it is that is appropriate. And I think that we've reached an exciting point in the development of the governance landscape, coming back to your question, where we're seeing more groups kind of embrace this idea that we really need to think carefully about what sort of things are needed, which requires engaging with stakeholders and impacted user groups. I think that's one really important component. We have a second really important component of then trying to build systems that can satisfy those needs, where I spend quite a lot of my time, and a third really important component is how regulators, policymakers and other governance mechanisms are able to actually monitor and enforce those requirements. All of these need to work, all of these different components need to work together in order to try to build a governance system, which is legally sound and technically feasible. And it's challenging, but I think it's really exciting and there's a lot of good work to do.

Ray Eitel-Porter [00:31:36] Maybe I would just add to that, Adrian, I think from a corporate perspective and the key thing that I would like to see, and which is increasingly becoming apparent, is senior leaders in organizations recognizing this as a really important topic and appointing someone at board level, at C-suite level to actually own the question of safe and ethical responsible A.I. however, we may wish to call it and then to actively engage externally with the bodies, et cetera, that you've been talking about and think about how to take those learnings and really enforce them within their own organization, raise awareness because I think it's a mixture of awareness raising and training, et cetera, plus also some hard controls and processes which need to go with it. And I think one would hope that maybe organizations can actually perhaps be even ahead of regulation in starting to lead the way, taking some of these ideas and really implementing them internally even before it perhaps gets too to the stage of regulation or official guidance.

Dr. Adrian Weller [00:32:52] Absolutely. I completely agree with you that there's a real role for companies to try to be leading the way and help policymakers find a sensible path forward that's going to work for them and achieve what society wants. So definitely companies should be part of the conversation and very much should come forward. I'll just give a quick plug that we, at the Turing, we've just recently started a trustworthy A.I. forum where we invite companies to come and talk about some of their issues with each other and with some academics and try to have a friendly Chatham House discussion to try to make progress on these topics. And we're grateful to you Ray. We know you're going to be helping to organize one of these. Looking forward to it.

Ray Eitel-Porter [00:33:40] Adrian, thank you very much indeed for joining me today. I wish we had longer to continue the conversation because I certainly found that fascinating, and I think we'll probably continue this debate and discussion both in person and over social media as we go forwards. I'd also like to thank our listeners for joining us. I think if I had to take three key things away from this discussion today, the first would be the topic of what do we really mean by fairness. The second would be the question around regulation, and do organizations need external controls or internal guidelines and how should we take that discussion forwards? And, the third point, I think, was really encouraging companies and individuals to get engaged in the debate and really try to help shape what is the correct approach for their industry and the particular applications that they are involved with. I would encourage all of our listeners, if you're interested in getting involved, please do reach out to myself and Adrian. We would love to engage with you going forward. And if you've enjoyed this podcast, please do subscribe to us through any of your usual podcast channels. We would look forward to you listening to us again. Thanks again for joining and thank you, Adrian.

Dr. Adrian Weller [00:35:02] Thank you, Ray. It was a pleasure.