AI and You

Transcript

Guest: Eleanor Drage

Episode 203

First Aired: Monday, May 6, 2024

Hello, this is episode 203! I am talking with Eleanor Drage, who is a Senior Research Fellow at The Leverhulme Centre for the Future of Intelligence at the University of Cambridge and was named in the Top 100 Brilliant Women in AI Ethics of 2022. She is the co-host of the Good Robot Podcast, "Where technology meets feminism," and co-editor of a recent book of the same name.

Last week we talked about her research and work on the podcast, plus some quantum mechanics, saunas, ham, lesbian bacteria, and… well, you had to be there. Catch up now if you need to.

Back to the interview with Eleanor Drage.

I've got a difficult question brewing here. Let me lead into it. AI has kind of jumped into something which it wasn't prepared for. I mean, when I took computer science, half the classes were mathematics. And we're talking about manipulation of symbols and something that's as far from sociology as you can imagine, which is probably the reason I got into it. And now, by virtue of the way that, in particular, large language models have worked, it's landed feet-first in areas that the province of human beings trying to work out social issues. My terminology on this is not as good as yours, so correct me as much as necessary. And in these cases, no one could claim that human beings have worked this out. It is an area of intense conversation, shall we say, and conflict between people, irrespective of any role of technology. So there are a couple of questions here that are equally difficult. One is, are we expecting too much of the technology at this moment? And the other is, is technology doing a better or worse job than people are at the moment?

I'm really glad you asked those questions. I actually haven't been asked them in a while. I always think that this, "Oh, is technology more or less biased than humans?" I think - this is not about you - but it's a lazy question, because we have to be ambitious for ourselves and of technology. Just kind of shrugging our shoulders and saying, oh, well humans are biased, too, is not the answer. Because we have to be creating things that are good. Take self-driving cars. Humans make a ton of accidents on the road. We all know this. I've done my driver awareness safety course. And yet we're not saying, oh, it's fine. Teslas can kill, like, the same amount of people per year that humans can. No. I think we know intuitively why we're not asking that question.

Right. And to be clear, I was not suggesting that we give technology a break there or don't expect that it should get better. But is there an undercurrent to the conversation that says technology is not doing as well as people are in this respect? By what standard are we judging it?

Yeah, well, I think we should also remember that technology is human, too. So technology is made up of human data, the traces of our lives and experiences, human data annotators and labelers. It's got its own politics of the organization that's building and applying it. So they're not distinct entities. They are and improving technology is improving society. And sometimes we think to make tech less biased, we can just fix it with code. But often we can't. Often if the police is deploying a tool and they're using it to shut down Black Lives Matter protests, it's not a question of de-biasing the technology. It's a question of companies saying, OK, should we actually not be selling this to the police to use in this way? It's those kinds of decisions. Right. So I think we must understand the human element of technology.

Right. I think it's our relationship with technology is the important factor here. I like to say that we shouldn't worry so much about the technology. It will keep evolving whether we do, whether we pay much conscious attention to it or not. The task is to build a better person. And the way we see ourselves reflected in technology might be instructive in that respect. Is technology reflecting to us useful insights about who we are?

Yeah, it can do. I actually thought that I'd encounter more AI tools that helped make the world better by showing us who we were. And actually, I have few, I see few examples of that being made. But one of them that I think is great is used by the FT to help put out job ads that are less gendered. And it's a tool that when you write a job spec, it tells you whether it's going to be more or less appealing to men or women. And it just points out for you, OK, we're saying like, we're looking for a bold, courageous leader. It's more likely to be a guy that picks up on that than a woman. And so those are really useful ways of helping signal stuff to us. They're not replacing human recruiters. They're not replacing candidates. They're supporting good hiring practices. And I think this is key, right, is that we're not trying to create tools that replace entire jobs, but support existing best practices and that align with those non-technical best practices. And the problems start when we start replacing best practices at work with an AI tool that is actually trying to do something completely different. Like, for example, video hiring tools that claim to be able to de-bias the workforce by removing race and gender from a candidate. Like, no recruiter tries to remove race and gender from a candidate. You're looking at someone. You can't do that. You know that they are who they are. And they're sat in the sofa in front of you full of their own experiences. And that's what they bring to the table, right? But AI is trying to do something different there, where they're trying to strip part of you from the way that you're understood by the recruiter. And that's just a terrible idea. It doesn't work, as we proved in the paper that we wrote that was published with the BBC and stuff. So yeah, it's got to align with best practices to be successful.

And you showed in that research that the interpretation of the video wasn't even looking at things that we would think to be relevant. Like, it could be influenced by the brightness and contrast of the video as well, I believe. But is this not because we're asking the wrong question? Like, the people have instructed this to say remove gender because we want to be able to say that we haven't considered gender, we couldn't be biased by gender. I'm reminded of, I think it was in Boston that started it, the symphony orchestra switched to blind auditions where they couldn't see the musicians. And the number of female musicians that they hired shot up as a result, which suggests that there was bias - even unconscious - when they knew the gender that was removed when they didn't. And people generally thought that to be a good thing. They even had to go to the extent of having auditioners remove their shoes because even the sound of those could provide clues. I mean, I think people may have been thinking of that kind of example when they were saying, let's de-gender applicants. Where did they go wrong?

Yeah, that's such a great example of the Boston orchestra, partly because it doesn't matter so much whether it's unconscious or conscious bias, because ultimately women were still not being selected. And although it does work in the context of the orchestra, and it's a good practice that I know is being done elsewhere. It doesn't resolve the issue of women being treated badly when they get to the orchestra. If there's an environment that wasn't there to encourage women from entering, it's probably not going to be great for the women once they are there. So being in an orchestra is tough as a woman anyway, because most of the music is performed at night. Rehearsals are notoriously sporadic, long. What do you do about childcare? Requires a lot of camaraderie. There's a lot of infighting and competition over whether you're the first or the second violin. So I think that that's a good analogy for AI as well. An AI tool may say, okay, I can get you more black applicants through to the next round. But that's meaningless. Because as we see in Cambridge when we now get more black students as part of our community, they might still have a really shit time because we still have professors in different colleges who think that black people are less intelligent, or who treat them in a really unkind way because they don't have the same accent or the same points of reference. So we need to do more. And any good recruiter knows that advertising a job to somebody who might be mistreated when they get there is a really sticky thing to do. We need a huge cultural change. That's hard. It involves incremental change. It involves very difficult conversations. It means pay equality, it means childcare, all of these things. So you can't solve a hard question with a simple technology. You mentioned the students that I worked with, the engineers on the hiring tool. And they were so lovely. And I felt so encouraged that this next generation of computer scientists were actually going to be involved in social technical issues because they spend so much time on Instagram and thinking about what's going on in the world. You'd hope there'd be an upside of that too. So they were all really aware. And we're going to really need that in the future. Because if you look at Cambridge Analytica, that scandal where Facebook and Twitter were giving away data to this company so they could manipulate election results by getting people to vote in different directions. The only reason why we know about this is because there was a data scientist there who was uncomfortable with what was going on. So we need to enable data scientists to have that inner voice saying, am I sure about this thing that I'm making? Actually, is this is this a good thing or not? Rather than just assuming that that's not their job, or feeling disempowered or unable to find the words, or this is just math, so what does it matter, really?

Right. And I think a lot of that culture was created by toxic masculinity in technology companies, because it was an opportunity for people with that kind of personality rising to the top to exercise that level of domination. And technology sort of broadly is like an ultimate male game. Create new toys, use them to win games and rule the world. And more thoughtful conversations are needed in all of these arenas. And this is the cause that you're promoting and as you're demonstrating, it's really complex and difficult.

Yeah. Toxic masculinity is bad for men, too.

Oh, yeah.

I mean, there's lots of guys who were building these frontier systems in big tech who really disliked the competitive attitude, the really horrible work culture. It's just really uncomfortable for everyone, really. So that's why we need to change these things so that everyone can be more themselves and exist in a work culture that's friendly and comfortable and familiar to them while also encouraging their best work. Interestingly enough, though, Margaret Mitchell, who was fired from Google for this paper about why large language models could be dangerous if they were too big. It's called Stochastic Parrots. And she says that one of the reasons that she was fired wasn't just the paper, which, to be honest, is kind of, it's a pretty mundane argument. We know that if you can't assess the data provenance, if you don't know where all the data came from, if you don't know lots of things about the data, then actually these systems can be dangerous. Pretty obvious, basically. What she was saying was because she was working with frontier systems, there was a lot of competition about who would be at the front line. And actually, that's why she was pushed out. And it's pretty sad because she's a great engineer, and now she has to spend her whole time talking about, as you call it, sociological issues, which is a waste of her time as well, because she should be spending time doing what she loves, which is building stuff.

You've mentioned Elon Musk a few times. He seems to come in for a lot of attention on your podcast, or the episodes that I've encountered. Does he hold some particular place for you in terms of his impact on the world, and what it reveals?

Yeah, I think I think we probably over-refer to these people, but at least everyone knows what we're talking about. It's like when we critique the way that AI is narrativized in film using *The Terminator*. There's lots of movies that do that in a similar way, but everybody knows what we're talking about. I first, I started to think more closely about Musk when I was reading his reading lists of his top AI-related science fiction books, and they were all quite similar, Asimov, golden era science fiction, and I just come out of this PhD looking at hundreds of years of what can be described as speculative or science fiction, and thinking how weird it was that all his favorites were from the same era and the same genre of writing. There's loads of really cool stuff out there, and there's loads of stuff that's not being made into real technologies, so we need to kind of broaden what these guys are reading. I'd love to be at dinner with him just to give him some books, give him some different stuff to flick through.

I know he's a fan of the *Culture* series by Ian Banks, which is a bit more modern. I've actually tried to get into that several times and couldn't. I'm actually more of a fan of the same golden era things, and it may just be that's what I was reading when I grew up, also because of being neuro-atypical.

Well, I mean, there's loads of really interesting books written by neurodivergent people and about neurodivergent characters, and certainly in our center we have like a very wide range of people who are interested in different kind of novels. So yeah, I can give you some books to read and for the reading list for anyone who's interested.

Speaking of science fiction, it comes in for a lot of attention in the podcast, it seems. That is a community that embraces causes like feminism quite tightly. If you go to any convention, you'll see that these issues are very much at the forefront of what's being discussed and presented. Is that something that you have engaged with in any way that is revealing to you about the role of science fiction in our relationship with technology?

Absolutely. It's why I got into science fiction in the first place. And this is before my hair was pink. So I'd go to these conventions, I think I looked relatively normal, and was so encouraged by people like quadriplegics being wheeled into different science fiction seminars, and how many wheelchairs there were at the entrance, people dressed up as different characters, people who looked kind of male passing and like long, thigh high boots. And I just thought, what a place to be alive and to be yourself. Because so much of the time, people seek this kind of escape from reality, because reality sucks for them. And I loved being around people who are imagining different kinds of worlds This is a sort of utopianism that's not based on necessarily on blueprints, on the kind of Thomas More version of Utopia, where he lays out exactly what the economic structure is. And you know, here's how society should be organized, because that kind of thing veers towards fascism. But instead, it's more about utopia as process, as thinking expansively, as thinking together in solidarity, just coming together and imagining things that are a bit weird and really accepting. And I think that society should be more like science fiction fandom.

Not for the first time on this podcast I mention the bumper sticker, "Reality is for people who can't handle science fiction." I was at the Los Angeles Science Fiction Convention last year speaking there. So everything you're describing is very fresh for me. And you've also talked about different science fiction movies at length. One of the ones I don't think I heard you mention - but I didn't listen to everything - was the movie *Her*. And I'm wondering what reactions you have to that.

What do you think?

I don't want to go there, actually. Because I can imagine some opposite kinds of reactions. I'm really curious.

Well, the first thing is, I actually am not a bad cinema companion. I think people always expect me to like come out of this stuff, like with *Dune* or *Blade Runner* and be full of critiques. But at the end of my day, I just want to go and watch something on a massive screen. And you know, but I guess with my academic hat on, it was an interesting exploration of, of the gendering of these kinds of social robots. It would have been nice if that critical edge had been a bit more apparent. But to be honest, my main critique is that we know that this is a dystopia. And we know that women are being sexualized as these kinds of romantic social robots. But you leave the cinema particularly as a woman feeling a bit dispirited. And what I want people to see is more of the like richness that science fiction offers. What about Aliette de Baudel's worlds where it's a Vietnamese future empire where women give birth to these ancestor spirit AI mind ships, where these ancestors, these ancestors are both scientific innovations, and in keeping with spiritual tradition, there's just loads of really cool ways that we can be encouraging people to think differently about the future. And I just don't think any of these films are particularly interesting, I guess, like *Ex Machina*, where you've got Nathan Bateman making Alicia Vikander shaped sexbots in the basement. You know, like, I think Hollywood knows that it can bet on these dystopias because people will, there's a, there's a, there's a fetishization of, of the dystopia. But I don't think that it's a socially responsible kind of film.

Interesting. Thank you. After you started the podcast, large language models hit the world in a big way with ChatGPT, that's when it became part of the public conversation in a way that never happened before. How did that event land for you in terms of your work? Was there a big shift?

There was a big shift. And I think to be honest, it was great because it increased public awareness of AI. And it got the conversation going. I see, I think the nice thing about academia is you can, you see in the long term, you're looking at how thought progresses over decades, rather than having to be a journalist and be sort of located in that moment, or you don't have to be so reactionary. I was really there was many interesting things about how the debate went. Firstly, I was thinking back to 1952, when the Ferranti Mark I computer in Manchester delivered its first AI-generated or sort of proto AI-generated love letters, programmed by Christopher Strachey, and directed to Alan Turing. And there are these great, quite queer love poems that do what GPT does, but in a way that was a little bit more experimental, and of course, really wasn't known about. This was an era when you only had a couple of dozen people interacting with one of these massive machines. And now everyone was sending me love poetry. And I was on dating apps at the time, when GPT came out, and I just got a sense of influx of these like absolutely appalling GPT generated AI poems.

Wow. With the book just out, what sort of reaction has it been getting? And I just want to mention how you said that you're, you were so excited about your book that your mum thought you were pregnant, and you had to let her down. Aside from that.

Yeah, we've done really well. So we've sold 300 copies in three weeks, which is good for an academic book. And we're doing our book tour, we've done a couple weeks ago, we did UCL DeepMind in London, and we're going to the US next week and doing NYU, Columbia, Barnard, and Washington. And the reason why I love this book is because it takes the people that are actually hardest to read. So the people I did my PhD on and puts them in conversation with the leading thinkers in technology, people who are at Google, like Blaise Aguariadkas, who leads the machine vision teams, and also the people who've been fired from Google. So obviously, these people aren't exactly the best of friends. And it just it's all these kind of different conversations around what good technology is. And I break them down and make them really easy to read. I promise you, I did the hard thing. I've made sure that that this will be the easiest thing to read of all these people that put stuff out there in the world. So this is your chance to read the most [readable] version of the most profound thinkers of our time.

And that's no small statement. Based on a world where the one of the ethics is, it was hard to write, it should be hard to understand. But as you said, you were removing the "-ologies" from the writing to replace with something that was more accessible. Just the context of the book and the podcast, *The Good Robot*: what is a good robot?

It's a title that was quite pithy and a bit silly and a provocation, right? I don't think anyone really thinks that there will be a good robot or that we represent that. But asking people the question, "What is good technology?" continues to be really hard to answer. And it's been interesting watching these like great minds really struggle with that response. And my favorite answers actually are not the ones that are really abstract and massive, like, oh, good technology is… some sort of philosophical response. But ones that are really embedded in our day to day reality, like Laura Folano talks about good technology being her blood glucose monitor. But only when it works well and the cap is on and it's not broken because those things really matter that actually sustains her. But equally, these devices for diabetics are made by companies that also create tools that help people monitor their calories and are often misused by anorexics. So there's real complexities around the technologies that people find good and when they're good and when they're not good. So it gets people to think about some of these headaches that they experience. Both really loving something and enjoying the magic of a new technology, I think is really important. I think that we shouldn't disenchant ourselves too much. We really need to enjoy things, but also in a critical way where we know what their effects are on the environment or on the labor force.

And we had another robot author on the podcast recently that you may know, Eve Herold, who just wrote *Robots and the People Who Love Them*. And there's a lot of crossover between what we're talking about here and that interview as well. Are we going in the right direction with the way technology is affecting these issues? Kind of a deep question to perhaps end on, but I mean, I'm thinking back to beginnings of the technology or computer era, and computers were women. And that's what the term meant at that time: Women sitting down with pencil and paper and working very hard at things that for some reason, probably men didn't have as much patience for. That changed. And then we went through a period of much more underrepresentation of women in computing. Now it seems to have elevated some. Do you think we're heading in the right direction or are there trends that concern you?

I'm glad you brought up the computers because my grandmother actually was a computer in Bletchley Park. But the only reason that women were computers was because it was deemed secretarial work. It was sort of pure number crunching. So it's that relationship between the value of the work and whether or not women are allowed to partake in it has gone really unchanged. In fact, I think it's got worse in the AI labor force. And Mar Hicks, by the way, she has this great book on the history of women in computation about how when women left the labor force, the UK computing lost its edge. And there's many different reasons for this, actually. My family had an engineering firm called Ferranti and the history of women in computing and the history of the company itself is really interesting on how these British companies failed, why they ceased to be competitive and why Silicon Valley went ahead. Anyway, that's by the way. I think that it's going in many different directions and I would urge people to take solace or encouragement or inspiration from organizations like Masikane, which is a non-profit that's trying to improve the representation of African languages in natural language processing. So in kind of the technologies like GPT, there's over 2000 African languages that are spoken by many, many people. These are just the most widely spoken languages. And the way that they're operating is interesting because they're trying to lean in to be included in these systems. Meanwhile, there's lots of activists looking at what the repercussions are of African languages being merely included rather than directing the future of these technologies. There's also AI being used to communicate with whales and a bunch of other kind of really interesting ways of imagining what AI would look like in the future with different materials that's not made of silicon and lithium, that's made of other things. So I think that it's moving in many different directions. And while the media focuses on all the really bad stuff, there's lots of cool things out there that are not really spoken about. And those will be the subject of Kerry and I's next book, Feminist Visions.

When's that coming out?

2025, I think.

Ah, when I finish a book, I need years before I can even lose enough of my memory to think about writing the next one. So thank you. For people who want to follow more of what you're doing, we'll have a link to the book in the notes and transcript. Tell them where they can find your podcast and what sort of audience it's for and what they can expect in future shows.

So, well, it's aimed at you if you're listening and if you're listening to this, you've made it the way through. It's quite a jovial show. It's not much of me. I just ask the questions and we interview people we think are cool. So they could be artists, activists. Forthcoming are the founders of Apple Together, the labor movement that was the first tech union out of Apple and just like an extraordinary, incredibly eloquent computer scientists. And God, we record so many episodes and batches along the line and you can find it at, will you link it in the show notes? Yeah, great. So just anywhere, Spotify. We also filmed video on YouTube, which is quite a nice vibe, me and Kerry in this extremely small podcast booth, usually dying of overheating. Watch us online too.

New meaning to Hot Takes. Well, Eleanor Drage, thank you very much for coming on AI&U.

Thanks for having me.

And that is the end of the interview.

You know, I thought about the question of what books I like, and this discussion may lose me a few listeners here, but it was thought-provoking, and I realized how hard it is for me to explain these things. I promise there's an AI-related point at the end of this. So, for instance, I do not get into cyberpunk. Stevenson, Gibson, do not do anything for me. Sorry to tell you this. There's going to be more bad news coming. On the other hand... I was told about the Nexus trilogy by Ramez Naam some years ago and didn't read it precisely because it was cyberpunk. Finally did do that because it has a lot about AI in it, and loved it. It was unputdownable. What’s the difference between that and other cyberpunk for me? Can’t tell you. Likewise, in any ranking of topics that I might resonate with, contemporary novels about southern lawyers would be near the bottom, and yet I have loved everything John Grisham has written and would probably read anything he wrote on any topic, some of which I have read that are not about southern lawyers. Can I tell you why those work for me? No.

Likewise, police procedurals are not a big thing for me, and yet I have loved recently the Sidney Fitzpatrick series by Robin Burcell. And also nuclear post-apocalypse stories of survival don’t do much for me at all as a genre, and yet I have very much enjoyed the Victoria Emerson trilogy by John Gilstrap. None of these things that I’ve talked about have anything in common in terms of politics or anything else that I can think of. What I like in books - and I read a lot of them - doesn’t even necessarily align with the same author. For instance, *Dune*. Tried to read that when I was 14, couldn’t get into it. Some six years or so later, a friend said, what, you haven’t read *Dune*? Go and do that. So I did, and it was fantastic, so I was ready for it then. I wasn’t aligned with my earlier preferences, and I read *Dune* and all of the five sequels to that by Frank Herbert. Yet, unlike apparently John Grisham, it’s not aligned with the author. I don’t particularly care for other books by Frank Herbert. Neither is it aligned with the story universe itself, because I don’t care for the other books set in the *Dune* universe that were written much later. So, having doubtless alienated a bunch of listeners by now, this is a long way to make a small point about AI and explainability, which is an important topic we’ve talked about on the show before. I can’t even explain my own preferences here. I could give you a purely *de*scriptive explanation, which is to say that it would be made up intellectually from what I could think these things have in common, but it would not be *pre*scriptive in the sense that you could do anything useful with it to determine what else I might like. Anyway, none of those particular mentions of books and their authors there was in any way an endorsement of the authors, or their political stances, or anything like that, yadda yadda yadda.

In today’s news ripped from the headlines about AI, Waymo issued a recall for its own self-driving car software after two of its vehicles hit the same truck minutes apart. Minutes apart from one another, two Waymo cars came across the same tow truck that was pulling a pickup truck in Phoenix, Arizona. The pickup was being towed backwards and at an angle rather than being lined up straight behind the tow truck, according to Waymo’s blog. The pickup’s front end was partly in a turn lane next to the lane the tow truck was driving in. Both Waymo cars incorrectly interpreted what their cameras were seeing and, because of that, wrongly predicted the how the truck was going to move, Waymo said. After a first Waymo vehicle hit the pickup, the tow truck kept driving. A few minutes later, a second Waymo vehicle came across the truck and also hit the pickup. There were no riders in either of the Waymo robotaxis at the time. After discussions with the National Highway Traffic Safety Administration, Waymo determined that it should file a recall report, which is done when a company makes a safety-related change to vehicles on the road. Since Waymo does not sell vehicles to individuals or other companies, the software update was confined to its own fleet of self-driving Jaguar I-Paces.

The pace of applications to the show for people to be interviewed has picked up again, and I’m rejecting three quarters or more of them for reasons that I’ve given before, a while ago, that these are people, more commonly their public relations representatives, who want to talk about the business of AI, and usually talk about how their company is dominating some particular market with innovative etc, etc. And this is not, as you regular listeners will know, a show about business, or how to make business decisions where the fact that the technology is AI is incidental. It’s not my interest, and there are lots and lots of shows that do a much better job of that. That’s not to say that we won’t have guests on the show who are CEOs of some startup that’s doing something innovative, and we’ve had plenty of them on the show. But we want them to talk about what’s cool, and enlightening, and novel, and interesting about the technology, and/or its impact upon the world around us. Not, not how do you make a buck off this. So that’s my philosophy with respect to the guests that apply to be on the show. Next week, my guest will be Gary Bolles, author of “The Next Rules of Work” and Chair for the Future of Work at the Singularity University. 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

Get the book: **http://humancusp.com/book2**