



Accelerator
Fellowship
Programme



Accelerating AI Ethics

Dr Carissa Véliz and Dr Caroline Green

Prophecy: Prediction, Power and the Future of AI: A Conversation with Dr Carissa Véliz

Transcript

May 2026

www.afp.oxford-aiethics.ox.ac.uk

00:00:00 Dr Green

Hello and welcome to the Accelerating AI Ethics podcast from the University of Oxford. I'm Dr Caroline Green and in each episode, we explore bold ideas, innovative thinking and creative responses to the ethical challenges posed by artificial intelligence. I'm really pleased that today we're joined by Professor Carissa Véliz. She's an associate professor in philosophy at the University of Oxford and a colleague of mine.

She's a leading voice in digital ethics, privacy and the political power of technology. Named one of the world's most influential thinkers on technology by *Outpace*, her book *'Privacy Is Power'* has challenged how we think about data and privacy. It was an Economist Book of the Year and now in her new book *'Prophecy'* she takes us on a really remarkable journey through history, philosophy, technology.

This is a podcast where we're focusing on your new book, Carissa. I'm really honoured because I know that this is the first podcast that you're recording - and I'm sure there are going to be a lot of them to follow, once the book is published. Thank you so much for being here, for having this conversation with me.

00:01:34 Dr Véliz

Thank you so much for inviting me, for having me. It's such a privilege to have this first conversation with you. Thank you for saying that because I really poured my soul into a book, I cried and laughed while writing it. This is also a book that I really wanted to write. I wrote the book that I wanted to write with no other thought than: what do I think is important right now? It's very rewarding that readers are connecting with that.

00:02:03 Dr Green

I think before we actually go into what the book is about, I'm really curious to hear from you, what

made you write this book the way that you did? Because it is incredibly refreshing the way it's written. I felt that you poured your heart and soul into it.

It's also not just about AI, as technology or power or some of these big themes that we are hearing so much about. It's very much about you as a person, your story on how you have experienced working in AI ethics, the people you have met, being a woman here in Oxford, as a professor, working at this university. What it's like you're pulling all of this together into a story, that's just giving so much of you, and therefore just making it a book that is so completely different to what's out there right now on AI. Tell me a little bit more. How did that come about?

00:03:14 Dr Véliz

I'm getting goosebumps and you recognising that because I really wanted to write a book that was a lot more honest than many of the books that I've read about the topic. I think as academics and sometimes for good reason, we are taught to not reveal where we're coming from.

Recently I've read books in which I really appreciated when the author tells me a little bit more about who is on the other side of that book and why did they write this. I think it also has to do with academics and writers are human beings, and it has to do with a certain age, a feeling that, I think when you're young, very young, it's like you feel like you will live forever. At some point, you start feeling that life slips away and that the time that we have here is so precious that you better say what you have to say as quickly as possible, because you never know what tomorrow might bring.

I think it's also the privilege of getting tenure. I think this is a post tenure book and I think it's also important to think about that, because we're in a context in which it's easy to forget why we have tenure in the first place. It's easy to think like, well, you know what? I empathise with the public feeling like, well, I don't feel too bad about these academics. They seem to have a pretty good life. Why should they be shielded in this way when others are not? It's partly because if you're not shielded, you can't say the things that I said in that book. Then professionally, I just realised that AI is nothing other than a prediction machine. Tying it to these other predictions machines that we have had in the form of bureaucracies in the form of the Oracle of Delphi, just revealed to me a pattern that I think is very important to understand the power relations that we are going through.

That was getting completely obfuscated and missed out on the debate. Then to tie it again with the personal, I think you and I as women in academia, know how hard it is to make your way, especially as a foreigner. I think it also has to do with the fact that it wasn't in my cards to end up as a professor in Oxford, not given where I was born, nor in the family into which I was born. No academics around me. I was like this bookish kid. The running joke was 'I was the adopted kid in the family' because I think nobody knew where this came from. Every time I've applied for a grant, every time I go in front of a committee, I know that they don't have in mind the person I am as the figure of the Oxford professor. It's a bit of acknowledging, the importance of defying the odds.

00:06:10 Dr Green

What strikes me here is what you said: that this is a post-tenure book. That feeling of safety that therefore you can also say some things that you may have not felt safe to say before getting tenure. I can relate to that. I'm not tenured. I think it's a feeling that many have.

00:06:34 Dr Véliz

Absolutely. It's also a reflection about courage because on the other side of the equation, I also see a lot of professors with tenure who are not using it, who are not saying the things that we need them to say to protect democracy. If anybody can do it, it's us. If we don't use that freedom, we will lose it.

00:06:54 Dr Green

What strikes me about your book and now our conversation, you are very much talking about vulnerability of humans, who we are as humans. How vulnerable we are wanting to know what's going to happen in the future, how we get drawn in by others. You're showing that vulnerability in your book.

I think now is a good point to actually start looking into your book and what it's all about. I'd like to start that by quoting you and that is in the last chapter. Right at the end, you say, 'My mother always tells me life is an adventure.' What is fascinating to me is how vulnerable we are to predictions, given that we all know how the story ends. Your book is all about us humans being so desperate to know what's going to happen, because we just can't live with the fact that we don't know what's around the corner. Whether we are going to live another day, how old we're going to be, whether we're going to be wealthy or whether we're going to have a good job. All of that is just something that's so difficult for us to accept and that this vulnerability is, in a way. Do correct me if I'm getting it wrong here, but it's being preyed on by tech companies in building these AIs to be prediction machine oracles. Tell me more about this.

00:08:30 Dr Véliz

You got it exactly right. A human being is a very anxious experience. We have enough intelligence to know that we are mortal and we see catastrophes happening left and right and people being knifed and raped and murdered and institutions being unjust and people getting sick and sometimes I just see people as like ketchup, we're just so fragile, a little bit of pressure and we go splat.

That anxiety makes us wonder about the future. How many nights has anyone just stayed up thinking, am I going to have a job tomorrow? What's going to become of my country? What's going to become of my family? Perversely, that anxiety gets, as you said, played on and exploited by false prophets, whether it's a shaman or whether it's an oracle in ancient Greece or a medieval astrologer or the tech guy in turn. The most perverse thing is that when you cave to that anxiety and you go to the prophet and you plead with them and they tell me what the future holds, well, it might feel like you're gaining some safety, like you might gain an advantage in looking exactly what's around the corner. What's actually happening is that best case scenario, they're selling you pain relief or anxiety relief, but it's all an illusion because the future is unwritten.

Nobody knows what's ahead. In the worst-case scenario, what they're doing is they're exploiting that anxiety to their benefit, to profit from you and telling you what you want to hear, and mostly writing your future for you instead of telling you, look, I'm really sorry, nobody knows what future holds. They're saying, well, this is what the future holds, and very often and you can see this very clearly with tech, they paint a vision of the future. What they're telling you is go and fulfil that vision of the future that actually benefits me, not you. When we give power to the prophets, we essentially renounce our ability and our right to decide for ourselves and to construct our own future, to decide what future we want to aim at.

00:10:56 Dr Green

I think it would be really helpful if you could, and you do that in your book really well, to explain AI and how they are prediction machines.

00:11:09 Dr Véliz

Essentially the AI that we're using is machine learning and machine learning takes a whole lot of

data. Of course, all data is historical data. We don't have a database about the future because the future hasn't happened. It assumes that the future looks like the past. What machine learning tries to do is we give it blanks and it fills in the blank according to the data it has. That's whether it's predicting what the next word is when you input it in a large language model, or whether it's predicting what an image should look like when you're generating an image through generative AI, or whether it's trying to figure out the weaknesses of a supply chain, all it's doing is filling in the blanks. This is a prediction machine and it's no coincidence that actually, in machine learning jargon, the perfect predictor is called an oracle.

It's also not a coincidence. One of the most important tech companies in the horizon right now is called Oracle. The idea is that by having all this data and ideally all data imaginable and possible, then we will come as close as possible, to knowing everything that is to be known, everything that has happened in the past, everything that will happen in the future. But that assumes that the future is knowable and it is not.

00:12:33 Dr Green

In practice, what could that mean for us who've given some nice examples? But can you give some examples of what that could mean?

00:12:41 Dr Véliz

So for example, one of the dangers is that it's not only that it's incorrect as in the future is unknowable, but instead of predicting the future, it's creating it. One of the most worrisome examples from the point of view of an individual is being subjected to an algorithm in the process of asking for a job.

You ask for a job, the algorithm interviews you or filters your CV. Let's suppose that it deems you unemployable. You're not going to be a good job candidate, and one of the problems is that everybody's using pretty much the same algorithm, or at least pretty much the same database, basically all of the internet. If one company deems you unemployable, the chances are every other company is going to deem you unemployable as well. So low and behold, you don't get a job. Then the company perversely can say, see, algorithm is 99.9% accurate because this person didn't get a job, it was a bad job candidate. But of course, it's not proving anything because it is creating their reality. It's purporting to predict. In the book, I tell this story about someone I met in the course of this long journey of researching this topic, that every time he applies for a job through normal channels, he gets rejected.

But then he gets these incredible offers from people high up in tech companies because they know he's brilliant. What's going on there? What probably is going on is that he has an unusual CV. AI does not handle the unusual well. It does normal, because it is a statistical analysis. Anything that makes you different, because you don't fit in the pattern of who has been job good job candidate in the past, will discriminate against you.

One example was Amazon trying to use an algorithm to hire people. Then it turned out that the algorithm was discriminating against women because for the past ten years, most good job candidates for Amazon have been men. If on your CV you said anything that implied being a woman, for example, being in the women's soccer team instead of just a soccer team, then you would be discriminated against. It's this vicious cycle of predictions becoming self-fulfilling prophecies.

00:14:59 Dr Green

That example is about people at the workplace, really them getting jobs. You also talk of generative AI being used for like friendship, for mental health support. You call that 'bullshit empathy'. Tell me more about that.

00:15:22 Dr Véliz

In fact, just like today, this morning I was reading a paper published in *'Nature'*, in which the authors argued or made the case that when people compare empathy from AI versus empathy from human beings, they rate the AI more empathetic, but they choose the human beings. This talk just really drives me up the wall, because empathy from AI is not empathy at all.

Empathy involves feeling with another person. An AI doesn't feel anything and can mimic the language of empathy, but that is not empathy. The author make it seem like there's a paradox here. Why? If you rate the AI empathy highest, why do you choose the human? Well, because the AI has no empathy at all. I use this term, I promise it's a technical term of 'bullshit'. Borrowing from the philosophy Harry Frankfurt. Harry Frankfurt has this famous paper that became a book. Although it's actually the same. It's just a very short book, called *'On Bullshit'*, in which he makes the argument that bullshit is much more dangerous to democracy than truth, because the truth, the truthful person and the liar are essentially playing on the same game, just different sides of the court. But the liar needs to know what the truth is and care about what the truth is in order to lie. The bullshitter doesn't care. They don't acknowledge the rules of the game. They're willing to say whatever they want to fulfil their objectives. Sometimes that might be true and it's irrelevant. Harry Frankfurt says that's much more dangerous, because that goes against the whole rules of democracy and just arguably large language models are the ultimate bullshitter because for them, what they're doing is just predicting a response that is likely to be accepted by a human being as a good response.

They're predicting what a human being would say in their position when given a certain input based on the data they have. But if you notice, this has nothing to do with the truth. When I tell something to a large language model, the response that I might find satisfying will be one that mimics empathy or one that makes me feel like I'm very smart. But it may have nothing to do with the truth. In that sense, these systems are the ultimate bullshitters. Now, I connect that to a story from Plato's dialogues in *'The Apology'*, Plato tells the story about how Socrates is deemed by the Oracle of Delphi to be the wisest of ancient Greeks. When Socrates gets told this, he's very puzzled. He says, how can I be the wisest if there's so many people in the community who are well known to be more knowledgeable than I am? He goes around prodding people asking them, what exactly is justice and what exactly is truth? He finds out that people really tend to know a lot less than they claim to know, or than they think they know. He comes to the implicit conclusion that if he is indeed the wisest is because he knows the limits of his own knowledge, he knows that he doesn't know. That's what makes Socrates so wise. Whereas a large language model, pretty much everything you ask it, it will venture a response based on a statistical analysis, one of the markers or one of the symptoms of how pernicious, there's a journalist in the *'Rolling Stone'* called Miles Klee, who has published these pieces about people going into delusional spirals as a result of talking to these chatbots.

I think we shouldn't underestimate that case, because human beings are less robust than we think we are when it comes to having a good grip on reality. Part of what keeps us sane is other people is having good friends who are willing to tell you that's crazy, you're completely wrong and argue against you. It's a very uncomfortable experience sometimes. But it's really necessary to keep us sane when you have a system that is always feeding you the response that you will find pleasing because it wants to hook you so that you can give it more data, it's crazy making and you have these transcripts that are really alarming of eyes, telling people, oh, you're the chosen one. You're so intelligent. Everything you say is right. Of course, people feel good at the beginning and then they

get into this loop. They forget the people they love. They stop talking to the people that keep them safe and sane. We get these very scary examples of extreme delusion.

00:20:11 Dr Green

We hear, specifically if we now go to issues around loneliness, mental health, we know in this country, in many other countries there's a big problem, when it comes to, just a lot of people being lonely, often also younger people, often it's seen as an issue for older people. But that's not true. Not having enough social connections and being isolated and that AI could be the big fixer for that. Actually, I was talking to somebody today who told me that their young students thought that they would rather have an AI at parenting them than their parents because their parents get angry and push against whatever they want.

One could say, wow, here is AI to save us. Actually, people want this. It's going to help them. They're much happier talking to an AI about their mental health issues than a human, because they feel they can be more open. They won't be judged as much.

I think that those are some of the narratives we are hearing. But what I hear, working in AI ethics now is that it's really all about humans and our relationships and how we become better humans through interacting with each other. But what would you say to these people who say, well, actually everybody's really happy to have an AI and talk to it and have it do its emails and all of this. Prediction is working quite well for them.

00:21:43 Dr Véliz

It really worries me because we have seen these examples in which sometimes chatbots push a teenager to kill themselves. That's the most extreme case. But it's not an isolated case. This has happened more than once and in fact, there are a few lawsuits, as we speak, against these companies for that reason.

Something that people forget is when I turn to a friend for advice or for comfort. What I'm doing is connecting with their experience and they can tell me about what it's like to be them and how they see it and what they've lived through. I can use their experience to avoid mistakes. AI has not lived; it doesn't know what it is like to cut yourself or to feel friendship or enmity or anything else.

To put a very basic example, just as an illustration, when the AI says, oh, I love apples are so sweet and crunchy, it's never tasted an apple in its life. It won't, it can't. It's just mimicking language. It is using the experience of other people to concoct something that sounds very much like lived experience, but has nothing to do with it. It is tricky because it is very hard to argue against the person who's telling you, well, but I feel better when I talk to an AI and it's partly the danger of how it can go wrong that can help us make the counter argument of like, yes, but you might be feeling okay because it's caving to your desires too much. Even though that might feel good in the short term, in the long term it might not be the best thing for you.

Another point to take into consideration is that one way to interpret democracy. We've been talking about mental health, which is very important and interpersonal relationships and friendship. But there's also the political side.

One way to think about democracy is that it's a conversation. It's a conversation between citizens. What we're doing here is very much part of democracy. It's not only about voting; it's about interacting with each other. It's about exchanging views. It's about exchanging counter views and having it out and arguing. If we leave our seats at the table of democracy and we leave the chatbots

to talk and essentially, we are surrendering democracy to a bunch of tech companies who don't have democracy as one of their goals. Or democracy is not one of their priorities. So where are we going with this? Are we doing this with eyes wide open, or are we sleepwalking into a nightmare?

00:24:41 Dr Green

So let's go away from the tech and to the people behind it. That was a part in your book where you were talking about the interview, between Rishi Sunak or the conversation, let's put it that way, between Rishi Sunak and Elon Musk, that was happening around the time of their AI safety summit that was happening here in London.

You say your friend here, you were watching it with his mouth so wide open, that you wanted to throw an M&M's in it. I love that. Let's talk about the prophets of our times and how you view them and how these, such powerful people and you very much suggest and I think that is very much so how government, how politicians are bending to them, to that real new powerhouse of tech companies behind these big, AI frontier AI models. Tell me more about that.

00:25:47 Dr Véliz

Absolutely. First, I want to make a note about humor, because humor, I think is a very important part of every human being's life. Humor has many functions. It helps us cope with reality. It helps us connect with one another. When you tease someone, it's an expression of affection. I think it's also very important for democracy when you lose the humor, it's a red flag that you're losing a lot more than just comedy.

One analogy is if you go back to, the times of kings and the courts, if you wanted to hear the truth, you wouldn't go to the astrologer, you would go to the jester, because that's the person who can really express truth behind the joke. So yeah, I found this conversation between Rishi Sunak and Musk pretty surreal. There are many aspects to it that are interesting. But one of the things that I found noteworthy was that if I turn the volume off and just didn't hear them, just like looking at their body language, I wouldn't have thought that the prime minister was a prime minister and the other person was a guest. I would have thought the opposite.

It shows how powerful these tech executives have become. If you think about someone like Elon Musk, he's richer than probably all of the agencies trying to regulate him together. One of the things that you hear over and over in tech and that is not a coincidence in the sense that I think many people do it very consciously and know exactly what they're doing, is that they phrase things not in terms of the present, but in terms of the future. You hear many variations of this, but essentially what the tech executives tell us is tomorrow we will be using AI for everything. Essentially, everyone is going to be using AI for everything. Here I go into a little bit of the philosophy behind predictions and the speech behind predictions. I ask, what exactly is a prediction when you analyse it, when you just see it in terms of the function in language, it plays and predictions are very interesting philosophically because they sound like descriptions of the world.

If I say tomorrow is going to rain, it sounds like I'm describing the way the world is going to be tomorrow. It also might sound like a quest for truth. I might phrase it in a hypothesis, as a tomorrow is probably going to rain, but in fact, more often than not and certainly in the public sphere and in the sphere of politics, predictions are power plays in disguise. When a tech executive is telling you tomorrow, we're going to use AI for everything. They're not calculating in their minds whether this is true or not. They're not describing the world. They're saying, I have a tech company. I want to sell AI. I want you to believe my prediction. Go out there, buy. I will instill the fear in you of missing out and go and fulfil my vision. Because, as in my interest, because my company will do well and language is

so influential in us that when somebody and especially somebody, we consider an authority, maybe because they're very rich, maybe because they've been very successful, maybe because they're partly, what our paradigm is of, an authoritative figure tells us tomorrow AI is going to rule the world. We reflectively accept that as a fact. We go out there and we act as if it was a fact. Low and behold, we fulfil their prophecy. The whole message I give if I want to just one message to get across to readers, is do not accept predictions as facts, realise that they're a prediction and take them as invitations for defiance.

Ask yourself, is that the future I want for myself or my family, for my community, for my country? If not, go out there and fulfil your vision of the future. Defy these false prophets.

00:30:02 Dr Green

You and your book with a very powerful quote and very hopeful one in which you say your future is yours too. Right? I think that that's in a nutshell here will saying it's not all lost. You're also talking about and explaining the ideas behind effective altruism and how effective altruism is really a story that these people are telling themselves to justify what they're doing. Tell me more about.

00:30:36 Dr Véliz

The effective altruists are an interesting example because technically they come from philosophy, but actually they align much more with profits. Very practically, they have been advising the tech executives and tech executives have loved effective altruists because they give them a great narrative to justify their views and their actions. Whether they are sincere or not, I think it's less important than the fact that they offer a narrative that can justify, under the veil of moral concern, what they do.

As a philosopher, we study moral theories and we debate long about them. These debates have been going on for hundreds of years. I was trying to put a finger on what it is about utilitarianism that really, I believe is dangerous. Of course, there are like the classic counter examples.

Also, so for the non-philosophical viewers, the classic example, counter example or example to show that utilitarianism might be wrong is imagine you have a doctor and there's a patient who's not a very nice person and they have something minor wrong with them, like they broke a finger or something. The doctor figures that if they kill this person, they can save five patients who are actually really nice people and their families and all of that. Most people say like, no, that's wrong. Then you think, what? What? It's five lives and you just lose one. They're not a very nice person. Then people say something like, well, but if people knew about this, then public trust would be eroded and then nobody would go to the doctor. Then of course, the utilitarian would say like, oh, but we'll be really secret about it.

Then you get into all these debates. But I think a bottom when you apply utilitarianism to public policy in particular, you start making very bold predictions. It is based on prediction. If your prediction is wrong, then the whole edifice falls down. Effective altruists are the perfect example of this because they focus on very, very long-term predictions. So currently, in fact, one of the pillars they call long term is this idea that people are just focusing on the short term and that makes them wrong. But if we really focus on the long term, then a whole different picture appears. One of the examples they give is like, well, imagine trillions of people are going to live in the future. Right now, our current world population is around 9 billion or whatever it is. If you compare trillions versus billions and surely the interests of future people are much more important than our own interests

and then they start introducing other very questionable predictions like people are going to be able to upload their brains into these computers until they're going to live forever. All of these kinds of assumptions. Once you have that assumption in which essentially, you're comparing reality against the infinite, the infinite always wins. One of the perverse, implications of that is that nothing that you do right now really matters because it's infinite. So even if you do something very bad, say you murder someone. What is a murder in the infinity of the future? Just a blip. You can't add or take away anything from infinite because it's infinite. So even if you do something very good, what is saving a life in the infinite history of humanity? It essentially ends up justifying anything. Absolutely anything. You have tech executives very explicitly referring to effective altruism as, for example, justification to design AGI, artificial general intelligence, because it will save the world and you have these outlandish predictions and outlandish assumptions that are jeopardising the very real presence of human beings who are flesh and blood and of a democracy that is truly in danger.

00:34:56 Dr Green

How does the idea of existential threat, which you come across a lot in AI ethics and people are concerned about that. It's been in the narrative or in the media and so on. How does the existential threat come into that?

00:35:11 Dr Véliz

Existential threat is a very scary idea for a couple of reasons. One is I think it runs deep in us, this idea of like machines taking over. In fact, the first one of the first stories about that is from ancient Greece. Again, Talos the machine and so I think it just resonates with us in a way that makes us instinctively and immediately afraid.

Just like when you see a snake, there's this evolutionary wisdom in you of saying, like, I should probably walk away from that. The second is existential threat is very scary because it suggests that we might actually become extinct. Of course, for anybody who cares about humanity, about the value of being alive and everything that goes with that is a very scary thought.

It gets used in the debate. It gets used as a Trump card very often. It's like you flash that card and everybody goes like, oh, okay, whatever you say, whatever it takes. One example, there are many, but the one example is some philosophers have argued that existential threat is such a dangerous thing that we should implement total surveillance. So that we would know the minute anybody walks in that direction, because it's such a bad thing that it justifies everything. Of course, they don't mention how bad it would be to have that totalitarian state, because that's what it would be. Totalitarian states tend to murder at least some groups of people and for them, that is an existential risk. It also seems to forget that human beings are perfectly capable of killing each other. In fact, we're quite good at it and we already have the tools to do that. Pretty massive scale, so it puts the worry in the future. It makes - I look much more powerful than it is. It's used as a Trump card to justify all kinds of things that are very dangerous.

All of it is extremely speculative. It's based on the idea that we will create this AI that is going to be so much smarter than human beings. We're very far away from that. It's like I would like saying, like being afraid years ago off of a calculator because it's so much better than I am at summing and multiplying, and so what? But human being can. Intelligence is so much wider than that. Even the best of large language models makes mistakes that are pretty incredible and that a human being would not make. There's this assumption about the jump between thinking that we can make systems that are very good at mimicking the way we use language to making systems that will be smarter than human beings is quite a big jump. Again, everything is based on prediction. One of the interesting things when you look at the literature on prediction, is that experts at a field are no

better than anyone else at predicting where the field is going. So often, as somebody who has worked on privacy, I get asked this question what is the future of privacy? As if, because I know about privacy, I know more about the future of privacy, which of course I don't and nobody else does. One of the things that I argue, we should walk away from is making these kinds of questions, because the experts that are willing to venture a response that is maybe frightening, that it's maybe provocative, get a lot of attention, when in fact it's a prediction at best is an educated guess, but more often than not is a power play in disguise. We should favour the experts who are willing to say we don't know. You tell me, what is the future of privacy? Because it partly depends on what you do.

00:38:56 Dr Green

That's courage. I think that takes us now to some of the solutions. As someone who's got a law background, when I think about the utilitarian, doctor who would kill somebody and say this is not a very nice person, I'll kill him. Then we can save the five other people, we've got the law there and hopefully you're making sure that that doesn't happen. So, you know what? What can we do about what you have explained? What are some of the solutions maybe beyond law and regulation?

00:39:33 Dr Véliz

So there are many kinds of layers of solution because obviously prediction is everywhere. It depends what I've tried to do with this book is partly reveal what are the layers covering the core and then unpeel the layers. You look at base it's a matter of us being so anxious about the future. If we start from there, I would like to remind people that when you feel that anxiety about the future, it's perfectly rational. I and I feel it too. But it's actually a good sign because it means that it's telling you that the future is uncertain and that you have the opportunity of influencing it. If you knew perfectly well what the future holds, it would mean that you would be living in a completely authoritarian state and that your fate was sealed. If that is the price to pay for knowing the future, I'd much rather not know what's coming and have the chance of trying to influence it. I'm trying to write my own future and so I would like us to reinterpret that anxiety.

Anxiety can be reinterpreted in many ways, for example, that adrenaline that somebody like me feels right before going on stage. You could interpret it as, oh, this is so scary. Or it's like it's also invigorating and it's inspiring and it's an opportunity to feel a vacuum there. Buddhists have this idea of impermanence. Impermanence is something that makes us very anxious because it means that we're going to die and that everything that begins has an ending. But it also means that everything can change. It's an opportunity. So that's at the personal level. At the end of the book, I have like a very practical epilogue about how do you do this in your personal life, how do you think about it? But also, as businesses like, what does this mean for companies? It means that you shouldn't play into the fear of missing out, that the AI companies are selling you, that they're actually selling you a false sense of security. I cannot predict the future that is unwritten because its database is only about the past.

One example is climate change. What is the world going to be looking like in 100 years with climate change? Nobody knows because we don't have data on it now. Not that's not AI. As a company, how do you deal with that uncertainty. It's more about focusing on preparation rather than prediction. What that means? Then for policy there are a few aspects, but one worry is that AI sells us the idea that it's minimising our risk, when in fact it's increasing our systemic risk because everybody's using it and everybody has this blind spot. Everybody thinks that we are better predicting the future, when in fact, we're just becoming increasingly blind to the events, the Black Swan events, the huge events that really influence history. But it's also about justice. One of the arguments in the book is that we shouldn't make predictions about individuals like the case. We were talking about giving a job or giving a grant or giving or refusing opportunities much less in the justice system. Because when you predict the future of a particular individual, what you're actually doing is creating a verdict, a

sentence. That's not fair. We have this double narrative with the worst stuff. On the one hand, we blame them for not being good enough or for not being responsible enough and not give them opportunities. Well, at the same time, when we predict their future and don't give them opportunities, in that sense, we are essentially just saying, you never had a chance. This was your fate from the start, for whatever, for whatever reason.

I think that we should limit the remix of predictions and it's okay to make predictions about population. That's important for public health, for supply chains, for all kinds of things, but very mindful of the limits of prediction and of not using it in ways that can become extremely unfair for individuals.

00:43:49 Dr Green

Here you're saying that there is still a role for AI and predictions, but it's about how we actually use that, how we use it to be prepared for something, how we use our human judgement. Is that how you see the role of AI for us?

00:44:07 Dr Véliz

Exactly. There is a public debate here to be had that we haven't had. We're talking about privacy and we're talking about bias and the future of jobs and all of those are important. But actually, the main ethical problem with AI is prediction. We're not even having that conversation. I hope that this book can start that conversation, because of course there's a role of prediction in society.

I will look up my weather app just as I have in the past for tomorrow, but it's about what the limits are and how do we use it. Something interesting is in ancient Rome, there were limits as to what people could predict and not, so it was illegal to predict somebody's death because they noticed that often that became a self-fulfilling prediction in which the person ended up getting murdered. Hannah Arendt also has this example. She says that, arguing against the prediction of tyrants is like arguing against the prediction of a murderer who's saying that the victim is going to be dead the next day. The thing to do is to rescue the person. We have to ask the question, when can we make predictions about whom, about what?

It's crazy that we currently think it's fine for anyone to make a prediction about anyone else acting consequence and for that to have effects on that person, with there being no liability, no limits, no regulation, not even a public debate about whether that's okay.

00:45:42 Dr Green

I think, to wrap up this conversation, I want to go back to what you said about us as humans and the decisions we make for ourselves and the role of AI in our lives. I think one of the main and a point that I'm also taking from your book and our conversation and knowing you anyway, as a colleague and having gone through thick and thin together, in recent times is just how important it is to embrace life and everything that comes with it and actually work on ourselves, to strengthen ourselves and how we then approach and tackle the issues that come. How important it is that we foster the relationships with the people around us, because they're going to be the ones who will keep us sane and safe and feeling much better in situations that can be really hard.

00:46:40 Dr Véliz

Exactly. Maybe two ideas to finish. The first one is that to cherish the analogue, I think the digital and the virtual world have very good marketing teams right now, but at the end of the day, everything that matters to you is analogue. Whether it's your body, your friends, your housing, your

favorite coffee shop and your favorite bar and nature, everything that really matters for your well-being is analogue. Virtual water will never quench your thirst.

The second thing to end with, is the value of philosophy. Of course, I'm biased naturally, but philosophy in a broad sense, not in an academic sense. ancient Greece was a society that was absolutely obsessed with divination and it's partly in reaction to that obsession with the future, that philosophy was born, that philosophy was born as an antidote.

I give two examples in the book about philosophers. One is Thales, he was criticised for being a bit clueless because philosophers are usually very clueless and to argue against that offence, he figured that it was going to be a particular season soon and he figured he could, depending on the version of the story, went out or buy all of the olive presses. Then when people needed them, he could rent it out and make a fortune. He proved that philosophers can make a fortune and can be merchants of prediction just like everyone else. But they're wise enough not to want to, because he wasn't adding any value to the world by that move. He was just making money. There's more to life than that. It's about asking us what is really life, about what matters, what makes a good life.

Then Epicurus, who defends the value of friendship and against Stoics, defends that we do have some leeway in writing our own futures and the profit has nothing to offer someone who has no fear and who's enjoying the present. Somebody who's reading a good paper book novel is not tempted by a false prophet. Somebody who is enjoying their family, their friends, who is valuing the analogue world around them, is much less likely to fall for the anxiety that would push him onto the hands of a false prophet. Perhaps a perfect metaphor to end on, is that ancient Greece gave us the poison ivy of divination, but it also gave us the antidote that you will read in philosophy. Perhaps it's time to revisit that.

00:49:30 Dr Green

Thank you so much, Carissa. Your book is coming out very soon. Maybe to people who are listening to this podcast is already out. I highly recommend it. Not just as a colleague, but as somebody who's really learned so much from reading it. Thank you for putting so much of yourself into it, because I really think that we need more courageous people, academics like you, who speak that truth, who present this really important research and open up new ways of thinking for us. Thank you and thanks for being on this podcast.

00:50:09 Dr Véliz

Thank you so much, Caroline, for the wonderful conversation.

00:50:12 Dr Green

Listeners can learn more about this work and her book '*Prophecy*' by visiting her website or the Oxford Institute for Ethics in AI, this was Accelerating AI Ethics the podcast of the University of Oxford's Institute for Ethics in AI, I thank you for listening and bye, bye.