AI and You

Transcript

Hello, and welcome to episode 111! Today's guest is Cansu Canca, calling in from Turkey, which means, if you're looking her up, that you replace the 'j' sounds in her name with the letter 'c'. She is a philosopher and AI ethics consultant, and founder of the AI Ethics Lab, prior to which she was on the full-time faculty at the University of Hong Kong, and an ethics researcher at the Harvard Law School, Harvard School of Public Health, Harvard Medical School, National University of Singapore, Osaka University, and the World Health Organization. She has given over 100 talks on AI ethics, including keynotes at Harvard Business School, the U.S. Department of Justice, and her TEDx talk *How to Solve AI's Ethical Puzzles*. She was listed among the "30 Influential Women Advancing AI in Boston" and the "100 Brilliant Women in AI Ethics." So if you've heard just about any of our shows, it's obvious why I'm delighted to have her here.

Why is ethics such a hot topic when it comes to AI? Ethics was not invented recently. Aren't its issues the same as they've always been? In one sense, yes. But AI is such a multiplying force that it amplifies the ethical issues of whatever it touches. Putting AI in the hands of a developer or a company uneducated in its use is like giving a toddler a chainsaw; they may end up cutting down a tree with it, but they're more likely to take their leg off. Or someone else's. And everyone we have on the show talking about AI and ethics goes in a completely new and fascinating direction. Why is that? I think it's because AI is amplifying human cognition at increasingly higher levels, that it touches everything we do and every way that we do it. Let's get into the interview with Cansu Canca.

Cansu, welcome to AI And You.

Thank you, Peter, nice to be here. Nice to meet you.

Thank you. And so, I'd like to know, as I ask most people, how did you get into this field? What was the trigger that said, "Oh, I want to work in the field of ethics in AI?"

I'm a philosopher by training and I have worked in population level bioethics, medical ethics, for many years. So, when I was teaching at the University of Hong Kong medical school, we were looking at a lot of these new medical technologies that are coming to the hospital, coming to the university and the conversation on ethics is still focused primarily and only on the physician's perspective, the patient's perspective, the healthcare systems perspective, the hospital authority, but it never actually goes into like, what goes on in the technology, or what happens when we are developing the technology? So, I started getting interested in the question of how do these systems make decisions? And what are the value judgments that go within those decisions that happen as the systems make these decisions. And this is 2016; it's not ancient times, but still a while ago, and 2016, really no one that I know of was talking about this thinking about? So, it was very like, okay, how do I go about this? So, I decided that I'll just take a little stop, break,

just learn about AI itself, learn about the technology, forget about the ethics for a little while, just learn how the system works, and then get back to the question of okay, then what are the ethical questions that I can see that happen as we are using these technologies? So, I started off with the focus on the health technologies. But once you deal with the technology and AI, you'll immediately get sucked into all these other questions, because there are so many interesting questions.

And your background at that point; you were already in philosophy: Did you have a PhD at that point?

At that time, I had finished my PhD, I had finished my postdoc, and I was holding a faculty position at the University of Hong Kong.

It's perhaps interesting, when I think about how many philosophers we've have had on the show, that maybe a lot of people would think that there shouldn't be that much intersection between computer technology and philosophy. But it is numerically huge. Can you perhaps speak to why there seems to be this big attraction between AI and philosophers right now?

I think it depends on how you look at the historical progress of this. In a way philosophers are always interested in weird questions. Like, how do you understand whether a machine is thinking or are they sentient beings, do they have a moral status? How should we treat a machine? These questions have been around like forever, nothing new over there. And questions about like, how we use machines, or how we interact with machines were there as well, but not specifically, they were not very pegged to the AI question. They were not so closely linked to AI. It was more like thought experiments and some actual technology that we were looking at. When I say we I mean the philosophers. And I think a good example of this is how the Applied Ethics especially and AI merge so quickly together is when you think about the question of self-driving cars, and how that picked up; almost immediately, the AI ethics questions started arising. The self-driving car question came up, like everybody, like there was all these, MIT had the Moral Machines Survey going on and people were talking about, well, how should the self-driving car make a decision and so on? Yes, but this was like a like this was the example was I taught the exact example the trolley problem was

Philippa Foot in 1967.

Philippa Foot, initially thinking about the distinction between killing and letting die and in the context of bioethics, actually, and then it was sort of dissected in many different ways by Thompson, then Frances Kamm. But the questions, the analysis was there, right? So, just to serve as a guide, moving it to the self-driving cars was almost like, well, okay, now here's an example that seems more realistic to you now, rather than a trolley that just got bent out of control. So, that I think is a good example to show like, when we had these questions, we were talking about these questions and they were very closely linked, it was just that the link was not there in real life. So, when the technology caught up in a way, with the philosophy, it was not that hard and also - I say this all the time - many of the AI questions, AI ethics questions, are not normal to AI. They are ethics questions, like resource allocation, or decision making, killing, and letting die, or

how do you judge human life and human value? These are the questions that we've been grappling with all the time in Applied AI Ethics so that you add some extra interesting questions.

And it seems to be very hot now in artificial intelligence and as you say, it's caught up, the technology has caught up, and maybe is that because the technology now looks like it's encroaching on the space of moral agency? And that we need to ask the questions that philosophers would pose of moral agents?

There is that question, but I think most of our AI ethics questions are not really on that. So, I agree with you, there is, as the technology progresses, the question of moral agency is becoming more and more relevant, because at some point, we are going to have a hard time understanding, what really, what is the morally relevant distinguishing factor between beings that have moral status and this thing that we created? So, this is a relevant question, and you cannot address it after the fact; you have to think about it beforehand. So, there is that track which is a very interesting question, philosophically a very interesting question, practically, it will be a very relevant question, if that if we get there. But a lot of the times, the questions are actually quite much more mundane. I think what makes AI so full of ethical questions is that we make a lot of decisions as we create AI systems. But in a way it seems like we never really paid so much attention about those value judgments. But now, these AI systems affect so many more people through so many different aspects of their lives, as they go and get their education, as they apply for loans, as they engage with the criminal justice system, as they go to the doctor. Now, they are so embedded into our lives, that their impact makes them something that we have to think about. So, in a way, a lot of AI questions are interesting because of their impact rather than their novelty. But getting back to your point, yes, there's definitely that super interesting novel, like really novel question.

Right. But, as you say that the bulk of the questions about ethics and AI at the moment are surrounding the fact that it's a technology that accelerates and amplifies the issues that we already had, but it just puts them on a scale that now we can't avoid them. Is that a fair way of framing it? These questions of bias and transparency; it just cranks that needle all the way up?

Yeah, I think so. Yeah, I think that's a good way of describing it. And also when you think about other hot like other areas, where we have a focus on, it's places like medical ethics, where you know who there is in medicine, that there is a set of physicians and there are a set of hospitals; whereas AI systems are developed in in ways that are not as controlled as medicine is developed or medicine practice. So, AI systems are both developed and practiced in a much more wide scale, but also in many different forms and coming from different sources, like it can be a kid with coding and then selling it to start making the startup and then selling the startup. It can be a big company, it just is. It has many different ways of entering to our lives.

I think that's a really interesting juxtaposition to make there between medicine and AI. Because one field is characterized by the ethic of f"irst do no harm," and the other is characterized by the ethic of "move fast and break things."

Yeah. We shouldn't say that, that ethic of move fast and break things, there's a bit of a problem with that one, ethically speaking, as you know.

And that brings me to the AI Ethics Lab, which you started. Can you tell me why and what it does?

Going back to that story in 2016, when I was in Hong Kong working on this question. I wanted to really focus on this, I wanted to take a break from medical ethics, from bioethics, and focus on the question of AI ethics. And prior to Hong Kong, I was doing my postdoctoral research at Harvard Law School. So, I figured, I'll just go back to Cambridge. I'm sure there are so many universities working on this question. If not, the companies are working on this, the tech companies must be working on this question. I'll just join one of the working groups I'll just get back to thinking and talking about AI, I find the group that thinks and talks about this, and I'll join them. So, that was my sort of thinking. That was apparently very wishful thinking; that was not the case at all in 2016. No, but I couldn't find anyone like, really. I talked to every name that you can think of in terms of academic institution and company. And yes, they recognized the issues, but they did not see it as like, here's this, here's a problem that has a name, here's an AI ethics problem that we should start solving in many different ways as we develop these systems, and as we employ these systems, and that was not really part of the conversation and moreover, whoever talked about it, they weren't that the conversation was very focused on either the technology or the legal regulation, the ethics, the what is right and what is wrong. How do we figure out what is the right thing to do? The real question, not compliance was not there. So, I had basically two options either I was just going to take continue with my job and wait for the field to - this is actually what I was told. They said "just take keep continue with your job, keep doing ethics, but keep an eye on AI so that when it starts happening, you'll be there to pick up the conversation." But I figured if there is no center, there is no platform where this is the focus, then well, let's tackle. So, I started the AI Ethics Lab, focusing on the AI ethics questions doing research on AI ethics questions with the intention to help practitioners solve their actual practical problems and this is something that I learned from my time in Harvard School of Public Health, where we had a great program in ethics and health. Again, this is like philosophers thinking about these questions and working directly with World Health Organization, directly with the hospitals, solving these questions. So, I figured that this must be the model, not just like, we sit here as philosophers think about it, and practitioners sit there too, and sometimes maybe our paths cross, but now the intention is, we will figure out, we will find the problems that you're struggling with and we are going to try solving them, so that we can continue the work of innovation together and better, hopefully, still fast.

You mentioned something there, which leads me to think about a distinction that I want to make in Applied Ethics, which is to a large extent, the interest of companies and ethics surrounds what's going to keep me out of jail. So it's about compliance. And I heard you make a distinction that suggests you're focused beyond that, which is what's right; never mind what's going to keep you out of jail. Can you talk about how that shows up for you and the work you do for your clients?

I think one thing that we should be always aware of is that truthfully, in ethics, we often don't know the answer until we work on it. It's not like "Oh, why didn't you do the right thing?" Well, it's often I mean, okay, there is a bulk of things that if you just mean to do well, you're likely going to do well. There is like a bulk of things that you can just get by like this. But any complex question, actually, we don't know the answer, we have to its sort of like doing rule of thumb math versus actual math. You know, we had a rule of thumb, okay, mostly get you by in your daily life, but if you're actually dealing with complex systems, or you want to do the proper math, so that your calculation actually holds. It's like that, like in ethics, we don't know the answer, we want to use our tools that we have from philosophy to work out, well, what is the right thing to do? And maybe it's not just one option, maybe there is like a set of justifiable actions that we can pick from, but there will be a set, the rest will be wrong, but we have to first determine that. And compliance sort of makes it simple thinking that we know the right thing, you just have to incentivize people to do that. That doesn't cover any of the complex issues. Yes, we still need the compliance, we still need the regulation going one step beyond compliance, we need the regulation, because let's face it, humans are not necessarily ethically motivated, let alone big companies. So, we know that we are not like, living in an imaginary world. But even in the case where you have really robust regulations, there are always grey areas, and the grey areas are very low practice most often. So, when you face a jury or a judge, or somebody who is accusing you of doing wrong, your argument will often be an ethical argument, not a legal argument, unless it is very clearly described in the law. So, ethics remains there prior to the regulation, but also after the regulation because of the grey areas and sometimes, because the regulation is just not ethical. That's not complicated.

And regulation is always going to be trailing what's necessary, they just can't move fast enough, particularly when it comes to technology. That's just a given. But with ethics, moral foundation, you have the chance of being out ahead of it, of how the situation is evolving. Could you tell me what would an encounter, real or hypothetical or idealized, between a client and the AI Ethics Lab look like as a kind of story? Like how take me through how they come in what they ask what you do, what they go away with?

So I have two jobs. I am running the AI ethics lab still. So, it's been over five years now. But I am also at the Northeastern University Institute for Experiential AI, I am the ethics lead there. So, the Institute is basically the same mentality as AI Ethics Lab. We are helping industry or practitioners from the academia solve their AI problems, in my case, the AI ethics problem. So, I deal with the responsible AI part of the work and same structure holds for both of these cases. Clients have questions. So usually, in AI ethics, because the field is so new, there don't come with very clear demand. But it's much more like, for example, "we have this project, and we are about to start it; we are worried about the ethical aspects." Or, this is a good scenario. The bad scenario, I mean, slightly worse scenario, is that "we finish this project, we are about to release it and now we are worried about the ethical aspects of it." That's like retrospective, it gets difficult because fixing something that's not done right is much harder than doing it right at the first time. So, this has been the typical engagement entry point, a specific project that sort of makes them

uncomfortable. Another thing that has recently started, which I'm very happy about after five years of trying to raise awareness and give talks, like all around the world, is that now companies, large companies, are coming and saying "We need to put in the whole structure so that we have an ethics component for all of our projects and this runs seamlessly and adequately alongside." And that's a much more complex engagement of course and then we are thinking about what do you have in terms of so I have a framework, a model that I developed, called "puzzle solving and ethics model," which I developed at AI Ethics Lab and now we are also using it at the at Northeastern University and here, basically, there are three main components of the strategy of a company. Do you have the playbook? Meaning, what is your guiding materials? Do you have use cases, do you have principles? Do you know how to use those principles? These have tools that will make it make these sorts of abstract ideas more available more easily usable for your developers, designers, non-philosophers, basically. That's your playbook. What is the guiding material you have? What did you develop? The other component is, how does this whole thing enter into your innovation process? Because this should not be like a review board that sits somewhere far away and just like says, yes, no, it should be a collaborative and a dynamic process so that we are not creating a bottleneck. But we are also not missing major ethical risks . And the final thing is, of course, training and finding and positioning the right people throughout the whole company to be responsible in engaging with ethics. That doesn't mean that we are turning everyone into ethics experts, which is impossible, and not desirable. But what we are trying to do is that there is a sort of like an efficient way for an ethical question to be considered, solved, if not solved, escalated to an expert within the company. So we don't want everything to go to an expert group. But we want to make sure that we train people, we position people, in such a way that they know what guideline to look at, they know which tool to use and it seems like it is a complex question they know who to call within the company. So, this is the bigger engagements that we are now finally doing, because finally now companies realize that there is a need for a more systematic approach.

And I'm curious how this breaks down in the category of the ethics? Could you, for instance, tell me like what sort of proportion of these cases do the ethics hinge on the curation of data that's used for training, or the foundation of an AI, and making sure that it is unbiased?

That's a very interesting question. You know, actually, even though bias something that we talk about so much, I think I have worked on only a couple of cases where bias was the main problem. It's very interesting; I didn't actually realize this until you asked, but the questions are really all over the spectrum. You know, sometimes they are about how you allow control for the user in what they are engaging with, or how you don't allow control or how you make others to be in control of other people using AI system, which is a problem, of course, we should fix. So, there are those questions. There are privacy questions that are again, not addressed legally but they can run throughout the whole process. The privacy question can come from how you collected the data all the way to, does this tool actually now allow people to access others' private information? It really goes throughout the whole spectrum. Or do you allow price controls for the user, if you're creating a product? So, it can be just a proxy question, did you realize that the proxy that you chose actually corresponds to a category that would make this product ethically unfair? So, did you read like, the simple example - I cannot give you the actual

example that we worked on - but the simple example is that everyone knows is if you use zip codes, you're going to be unfortunately have your system clustering according to race, right? So, because the zip code, and in the US, the zip code and the race are really pegged together. But there are many different ways that this happens, not just the zip codes. So, you have to be very careful that the proxy. So it's really all over the spectrum. Bias is not the only thing, neither the privacy even though the AI ethics has been very vocal on those two questions. For the time being. I think, sort of going through phases. You know, initially, it was like that, I think bias, privacy was a big thing and then we added the bias, hopefully, we are going to keep adding these topics so that people know about it and understand the importance.

That's the end of the first half of the interview. This is one of the ones we split into two so we keep these digestible. I thought that was interesting that Cansu came to AI ethics from medical ethics, because the moral imperatives of those fields are so radically different, as I said, and it makes me wonder what would be the result of bringing the medical ethics viewpoint, like the Hippocratic Oath, into AI?

I've got to mention, by the way, that Cansu and I met at an online event hosted by MKAI, which is the Milton Keynes AI community started by Richard Foster-Fletcher, and I am beyond impressed with how he's built up this powerhouse platform dedicated to making the AI area more diverse and inclusive, and populated it with all kinds of wonderful people that he's thoroughly empowered to take it to whole new levels. Go to mkai.org and check it out.

In today's news ripped from the headlines about AI, China plans to use AI-powered robots to build a dam – yes, a hydroelectric dam – in Tibet. I think China must be the world's leading consumer of concrete and they certainly have demonstrated the ability to build things at ridiculous speed, like at the beginning of the pandemic when they built two 1,000-patient hospitals in a week, and put webcams on the sites to show that. This dam will be built in two years and the robots will act as a giant 3D printer to build the 180-meter structure layer by layer. The Yangqu Hydropower Plant will be constructed over the Yellow River, generating electricity for 50 million people. It's a bit hard to figure out its output, because reporters confuse power and energy. A lot. But the reports say 5 billion kWh per year, which is an average of 570 MW, and other reports say it will have a capacity of 1,200 MW, and that sounds right, the peak generation rate seems to be about twice the average.

Next week, I'll conclude the interview with Cansu Canca, when we'll be talking about the ethical issues of search engines and recommender algorithms, and getting yet another take on the Blake Lemoine incident from an ethical viewpoint. 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.