TECH REIMAGINED
AN ENDAVA PODCAST

How can we ethically use data to put AI to good work?

00:00:12
Bradley Howard : Hello everyone. I'm Bradley Howard and welcome to a new episode of Tech Reimagine. We're now in season two, where it's all about the big questions around technology and the industries that impact our lives. Today's big question is around how we can use data ethically to put AI to good work. And I'm happy to have with us today Tom Gruber. Just to remind us everyone, Tom, would you mind saying a few words about yourself?

00:00:36
Tom Gruber: Yeah. Hi, Bradley. Thanks for having me on your show. I'm Tom Gruber. I was the Co- founder and Head of Design for the company that made Siri that we have on our iPhones, and I have a series of companies that involve AI over the years. And nowadays I'm doing a consultancy called Humanistic AI in which I'm advising mostly startups in how to use ethical use of AI.

00:01:00
Bradley Howard : Well can you share with us some of the ethical considerations and challenges about using data for machine learning and artificial intelligence.

00:01:08
Tom Gruber: Yeah. The big topic, let's start in on essentially why would you bother with this? It turns out that if you use a powerful technology with no ethical grounding whatsoever, you often end up with a Frankenstein story. That is the unintended consequences come back to bite you. And we've seen that, for example, in the untethered use of AI to optimize addiction on social media which is causing enormous social harm. And now Frankenstein's out of the lab and it's hard to come back. So the better approach is to approach it with eyes open, especially ethical eyes.
So what does that mean? It doesn't mean do good. Doing good is not, as they say, not operational. It doesn't tell you how to start. AI and ethics is really about what is it about the technology that gives us levers that we can make decisions or choices in how we use them, that will give us the kind of outcomes we want? So it's actually really a practical thing. So for instance, if we don't want Frankenstein don't make monsters that can walk outside the lab and scare people. Now let's talk about the basics nuts and bolts of this. Data drives machine learning. Everyone knows this.
So one of the core pieces of this is if you're going to make a model that is going to be used to train them, an agent or an assistant, some kind of AI thing, that's going to make decisions and that model's based on data and that data is supposed to represent something about your decisions like what you think the right answer should be, then everything about the data matters. Because you're basically what you're doing with the AI now is you're essentially crystallizing down the labeled correct answers that are in a data set and saying that's going to represent me, or my company and my government it's going represent me when these decisions are being made.
So a lot of people are looking at the problems of data. So what are some of them? The obvious easy ones are like, well, if the data doesn't include representative sample then it's the wrong data. We know that from medicine, from science in general. We don't want to build things that are trained on one, population multiplied to another. And it could be gender, age, race, all kinds of ways you can make a mistake there. That's been covered a lot and is kind of obvious, but it's easy to overlook that bias. That's so- called data

bias. There's another thing that's going on, is essentially the representativeness of the data. That is essentially a basic element of evidence- based reasoning.

And of course the drug administrations of the various countries like FDA in the U. S. that approve a drug, their whole reason to exist is because making judgments from statistical data sets about whether something works or not or whether it's appropriate for a population is a subtle and difficult decision it requires careful thought. So that means in AI, which is basically mostly these days a data- driven statistical kind of inference, we need to make sure that we have data that would in fact sanction such an inference. So for instance, if we said, " Oh. Well, I'm going to make a medical diagnosis system and I'm going to give it a tiny data set because that's what I had as a graduate student to work with it in college."

And then I say, "Now I'm going to sell that as a company to other companies to use it as a test." That will be unethical and stupid because you're going to be basically making mistakes. And people get away with it because it's very hard to evaluate them. Now I want to add one more element to this. And this is something that I think most of the conversation has overlooked. AI and machine learning isn't just garbage in garbage out. Meaning that the data is not the only thing that determines the bias or the outcome. The other thing that determines is something called the objective function. So an objective function is what happens when you train an AI model to say, " Okay. Classify images or make decisions."

You're training it against objective model that says, what does it mean to get a correct answer? And sometimes the objective function is just match the training set, but other times it's like optimize a goal. So for example in the social media case, the goal was do whatever it takes so that when you put stuff in front of humans to watch it makes them stay on the site and click on it. And the AI didn't know anything else except the objective was to do those things, and so it created this highly optimal system that addicted people. It wasn't that the data was bad, the data were just people using social media systems watching videos and so on, but it was because of the optimization function, the objective function it didn't account for all the stakeholders, put it that way.

So now if you think about that in terms of your own policies, if you build an AI whose objective function is aligned with human benefit you're much more likely to get an AI that has the intended consequences and not the unintended consequences. Because it will be drained and evaluated against objective functions that optimize for those goals.

00:06:15
Bradley Howard : Just playing devil's advocate on this. So is it possible for the social media companies to say, " We want to somewhat optimize this feed rather than make it highly addictive."

00:06:30
Tom Gruber: Oh, absolutely. I mean you could think of it as, I mentioned stakeholder, you can think of it as a multi- stakeholder problem. It's like analogous to an extraction industry like oil or something. If oil wants to be, big oil wants to be completely unregulated then all they have to do is take stuff out of the ground and throw it up in the atmosphere and cook the earth and not be accountable for the externality. Then they can do that. But if we let them do that, we're going to all die. So in the social media case the analogy of this is, we allow them to take the attention data from billions of users and use it to addict them and make money and that's all they do.

However, the objective function could have multiple stakeholders. One stakeholder is the corporate profit, the other stakeholder might be the mental health of the users. And if you actually account for mental health, wellness attention, whatever, there's a lot of ways of

measuring it, that will change the objective function. So let's take a simple example. You want someone to play a violent video game, they get into a trance. They're like, " Oh my God. Dah, dah, dah, dah." Right? There's no reason that the game console should be ignorant of the fact that that person is now in a trance. There's plenty of technology to know that that person has been coming into addictive flow state with the game, in fact that was what it was designed to do.

But in particular you can look at pupils, you can look at blood heart rate, you can look at the voice tone over the microphone. There's a million ways of detecting that this person is not well. They're not thriving as a human being right then. And so you could literally put in to the game another optimization goal that says, we want people to play the game and be happy and be well. And it would transform the dynamics of games into much more social interaction, play kind of environments and much more like individual addictive trances. And the same thing can be done with social media, video watching, a lot of these cases where we've over- optimized something.

00:08:23
Bradley Howard : So what's your view on the social media companies using AI to make their platforms much more addictive in the feed and nudges, et cetera, but not being able to use AI so effectively to remove a lot of the bad content? And I'm well aware that the social networks claim they remove millions of items per day, but it doesn't feel quite as effective as how addictive the platforms are for general use.

00:08:50
Tom Gruber: That's a good point, Bradley. I think that's the key. I think you nailed it. The key is they need to be accountable for the externalities of the misinformation. That is their platforms are being used deliberately by propagandists and by saboteurs to screw things up and using misinformation, right? And they're also accidentally pushing misinformation because the algorithms are blind to the content. The key is to have the algorithms not to be blind to content.

So an algorithm that recommends that you watch a video needs to know is this video a conspiracy theory video? Is this the video full of known falsity? Falsehoods? Is it run by an organization that's known to produce hate speech? These are things that can be known. Movement has been made in that direction. But the excuse that our AI isn't good enough is a silly excuse. AI can be made to be good at things people spend time and money making it good at. And these companies certainly have the time and people to do it. So I think the argument is not valid that they can't, it's just that they haven't been highly motivated to do so.

00:09:58
Bradley Howard : How do you think that companies can prevent bias appearing in AI results? There's been some high- profile ones, the Microsoft, Twitter bots, and even the Apple cause with its credit limits, et cetera, found a gender imbalance. How do you think companies can prevent that bias where the bias wasn't really programmed in it at the very beginning? In fact, one could argue the AI is completely impervious to its agenda at the onset.

00:10:28
Tom Gruber: Well there's lots of approaches to it. I mean one is actually bloody- minded, if you'll pardon my use of a (inaudible) expression, which is essentially to be like an affirmative action. Like essentially say, " We'll deliberately build into the objective function that the decisions made by this thing should not be gender biased." Now what you will

get is a suboptimal achievement of some other objective. So if the objective was match the training data with the highest possible accuracy and if that's the only objective, you end up with reflecting the biases of the training of sample. If you add these sort of the other goal of and make sure the decisions are neutral with respect to gender, you have to include a lot of data that are harder to acquire than the training sample.

Well in some cases knowing the gender is not a big deal, but knowing what cues are triggering the gender bias are tricky. But we've done this as a society. And at least in the U.S, there's been all kinds of laws passed and experiments done over the years about, for instance, truth and lending. When you lend someone money to buy a house, there are all these rules about what can or cannot be included. When you hire someone you're not supposed to ask them even their gender, and you're not supposed to ask them how many kids they have and where they live and all those sort of things. Because those introduced biases.

It makes it harder to know who the person is, but it also makes it in a sense fairer. Those experiments don't always work sometimes they're misguided, but those are the sort of things you can do. The thing about AI that makes a difference is that you can build into the objective function, literally a computable function that evaluates how you're doing on that bias score. And that's how you can make progress in this area.

00:12:11
Bradley Howard : Will that also help ensure the AI is protected against other attacks and abuses?

00:12:15
Tom Gruber: Yeah, I don't know. That's a hard problem. Essentially, AI being a computer technology is going to be vulnerable to cyber attack like other computer technologies, like database records of credit cards and so on. And so there's the fundamental problem of data security which is going to hit AI like everything else. I think what's potentially different about AI is that the ability to impersonate people, once you have this data the ability to impersonate them is being amplified. It's being made easy by a new AI technology. So the ability to commit fraud is going to be much cheaper and easier to do because of AI. And that's something we need to be careful of.

There's already an industry being built around counter measures. So we're having an adversarial war between the bad guys and the good guys of the use of AI. But it's really moving so fast that we're going to see massive amounts of fraud. And it's not clear whether these attacks will be something that overwhelms the e- commerce or the trust with pink companies online or not. We won the war with spam, but it wasn't one with AI in the same sense we're talking now. But it was a sense, a kind of fraud detection, so it is possible that we could do this. It's just going to take a lot of work.

Let me say one other thing is that the goal of protecting people against abuse is oftentimes the goal of educating people about how not to leave weak front doors open. Leaving things weak. And this is a thing where AI can actually help a lot. It should be able to be a protective layer around individuals so that they don't have to worry about data security, so it can be taken care of for them. So if grandma's about to respond to some spam or some fishing expedition and giveaway her bank account number, the thing can stop that from happening. Recognize that pattern to stop them. This is going to require the platform companies, the phone companies, and so on to deal with the problem and give this... Build those into the systems.

00:14:26

Bradley Howard : Linking your experience at Siri and some of the subjects we talked about today, can you see us using voice assistants to pay for right soon?

00:14:37
Tom Gruber: Well it's just an interface, so of course. In fact when we started Siri we had enormous number of business case, use cases. You can buy pizza, you can buy hotel rooms, you could buy movie tickets, restaurant reservations. And all these things, these e-commerce transactions were already enabled. So it's a natural act for virtual assistants. Absolutely. I think the question gets interesting when you take something like Alexa, which is of course attached to a company who is the world's leader at e- commerce. Is there a sense of a symbiotic relationship there and will that change the character? We haven't seen it happen yet, but it'll be interesting to see when that happens.

00:15:16
Bradley Howard : Yeah. And can you see a time when we'll start getting proactive alerts from voice assistants? At the moment they're very reactive. But can you see a time when it will start prompting us, " Oh, have you now run out of that item, that item? Do you want to just order some more?" You say yes and then the payment goes through and they fulfillment.

00:15:35
Tom Gruber: Of course. I mean, we built the Reminders app. We asked Apple to build the Reminders app for Apple because we needed that capability to complete the vision of Siri. And so that's the way the assistants do it today is they just hook into the notification systems, the reminder systems that are already in the platforms. And so Amazon has a renewal machinery that has a one- click purchase. They have all these ways you could be reminded, they just have to hook the user experience of the assistant up to that. There's a lot of easy examples like order somewhere paper it's not a big deal.
But think about the more subtle ones. How do you get reminded to take your medicine? It turns out medical and compliance are taking your medicine as prescribed is a massive problem for healthcare. Is a huge source of medical mistakes and costs and so on. What if AI assistants were in the business of helping you remember to take the right medicine at the right time? That could have extremely life- saving effects. But it's also a lot more than just turning on an alert, in this case it can be a conversation. Did you take the medicine? Which one did you take? How do I know? And so on. And that's where you needed a little more intelligence than just a proactive notification. And I'm looking forward to those models maturing and getting better at that.

00:16:58
Bradley Howard : Final question is on the technology platforms. So we find ourselves at the moment with only a handful of public cloud providers and they're offering their own AI tools. And they're also using common sets of learning data and training data. Do you think there's a risk of having two similar AI patterns for multiple services and that that might lead to some biases from data sets which a company might not have proactively given themselves? It might've already learned it beforehand.

00:17:34
Tom Gruber: I think it's true that the cloud providers can do certain kinds of inferences for higher kind of speaking. You want to classify images, give us some images from your domain and we'll classify them for you. Speech (inaudible) that way, there's several services like that. These days these big cloud providers, as you're saying, start from a

base model and then customize it for the customer's data. And yes, of course they could have biases. Now the interesting thing though is that the larger the platform company the more likely they're going to have a more representative, globally representative set of data. And also the more competitive they are then the more likely they can compete on the nature of their data being bias- free. That can be a feature that you can compete on. And so since these services today that the AI services that are provided by cloud providers are basically commodity inferences, they're commodity computational services then they are subject to the normal benefits of free market capitalism in which there's competition among competitors and the equality will rise from that. One thing that I am concerned though about is that there will be certain models that are so expensive to build that only a few cloud providers will have them. So we've seen, you've heard about these language models GPT-3 and Google has one called LaMDA now, which are astonishingly good at completing a thought, so to speak. A written thought. And you type in some words and then you say finish the story and it makes what looks like a human written story. It's astonishing how good that is.

Now, those are expensive to build. They're not expensive to run by the way, they're just expensive to build. And the problem here is that if only five companies on earth can build them, we're going to have a problem because those companies are going to create this weird Silicon Valley bubble to represent the rest of the world. It turns out that's not true today because the way those models are built is by basically scraping publicly available data off the web. And so they actually are free to and there's no reason not to go as broad and wide as possible in sampling the data. In the future though, there will be special versions of those kinds of services that are based on data that isn't freely available like healthcare data. And when that happens we have ethical decisions to think about, about who owns the data and who has the right to exploit it.

00:20:05
Bradley Howard : Definitely. On that very serious note, thank you very much for joining us today again, Tom, and the chance to pick your brain about the ethics of AI. What a great conversation. Hope that you enjoyed listening or watching this episode of Tech Reimagined with Tom Gruber. And thanks very much for joining us today. If you like today's topic, then please show us some love by hitting that like button. Please remember to subscribe and tune in next week for another episode of our podcast. If you've got any other questions, please visit endava. com and hit the contact us button on the right- hand side. I'm Bradley Howard. This has been Tech Reimagined. Until next time.