Podcast: Responsible AI – Mitigating Data Bias

00:00:02

Good day.

00:00:03

Welcome to another episode of The Elevate Podcast series. I'm your host, Andy Packham, Senior Vice President and Chief Architect at the Microsoft Ecosystem Unit at HCLTech. In the last few episodes, we've been focusing on practical advice on how AI can drive positive business outcomes at scale.

00:00:22

One theme that repeatedly surfaces in these conversations is being made of responsible AI and the importance of ethics. Not only in AI itself, but also in the underlying data which you know every I system relies on.

00:00:37

I think in many ways we really shouldn't need to be talking about responsible AI. Ideally, we're all going to be acting responsibly. We're all gonna be acting ethically and we're all gonna make sure that the IT systems that we provide and support act in the same way.

00:00:53

But the world is never clear cut. I look at this in two lenses. There are always going to be entities, organisations, people that don't act to that set of norms that we would generally define as ethical, but for this conversation I want to set that aside. And today what I want to focus on is the vast majority of organisations that want an aim to use AI ethically. But are finding execution difficult.

00:01:18

Our industry is based on trust and it's important that we focus on this complex subject. But despite all of the time and effort that we have been, I think organisations are still struggling to understand what really needs to be done.

00:01:32

To create AI models underlying data that you know reflects that important subject. So to unpick this today, I'm once again joined by two industry leaders to go through this subject and provide some use usable advice on this area.

00:01:53

I'd like to kick off by introducing Srini Kompella, Senior Vice President of our Digital Business Services at HCLTech and Jeeva AKR worldwide leader, Azure Cloud Scales, Analytics sales strategy and execution at Microsoft.

00:02:10

So let's right dive right into this. So it's really what drives bias in data and in AI algorithm.

00:02:19

Great. Andy, thank you so much for the opportunity again and I'm extremely pleased to have another opportunity to share my experience and also my point of view with all the HCLTech audience. So really pleased for the opportunity. Thanks again. First of all, great context that you have actually set it up for this conversation today, right. So if you think about where are the sources that are, where the data bias can actually come in?

00:02:47

Literally, there are about handful of them, probably 5 or 6. So very first one that anyone can understand right now is that as simple as the data bias right? And we all know this that any time when you are actually training data that is purely based on either on one particular population. Or one particular demography or any one particular set? Then the entire model is going to provide an output that is going to be based on the data that we have trained the model on and it is going to exclude the other ones. So of course data bias and it's also called the sampling bias as well, right? So if the training data is not representative of the entire population, of course, the AI models will generalize inaccurately. And then the second thing also is that overfitting to the training data? This is something that I see very commonly among the customers who have embraced the use cases.

00:03:49

What I see that is that when you are training an AI model, particularly on the specialized on the training data. Then when you are introducing a new data set, those models tend to perform very poorly because they have not seen the data. I think this is something that we have seen it as overfitting to training data. This is more about the data bias or sampling bias or overfitting to training data. This is more on the very basic level of the factors that can contribute to the data bias, but if you go up the chain, right? And this is where we need to get better at it. And this is where we need to focus on, which is about the implicit biases that are held by the designers in the prejudice and the design, right. So, for example, either knowingly or unknowingly, they can actually induce implicit biases that and actually transferred into the the system's behavior and that can actually be a big problem for all the models when you're executing it and then the temporal bias, right? So temporal bias is something that you train the data, train the AI model based on a data that is collected at a very specific period of time and it may not reflect the current reality, right? So for example, if you ask someone about if you train a model of all the historical events that happened in South Africa up until 1980s and if you ask someone a question is about is all them people from all demography are allowed into the restaurant.

00:05:24

Of course, the model is going to say that it no, it's not. And so that's clear temporal bias, right? But that's things have changed tremendously since then.

00:05:34

So one thing is that it is about when we are training the data. It is extremely important that we actually provide both historical context to the data and also the real data, right. And then the final piece of it is that.

00:05:48

It is the social, the social factors, because at the end of the day that AI models do not operate in isolation, right? And what we have seen is that you can actually train the model and you can actually bring the data and the AI models will provide the output. But most of the time.

00:06:08

What we are seeing is that they sometimes when you don't provide the context to which the decisions are being made, it can actually create bias as well. So you know, all I'm saying is that there are several countries contributing factors as simple as the data and the sampling by assets all the way to the practices in which the AI models are trained and my opinion is that I think addressing these biases require a very holistic approach. Andy, in my opinion and also that it should not be just based on the technical aspects and data aspects, but also the social, the social context to which these data is being trained on. All of these things are actually very, very important as we work towards building a more fair and also trustworthy AI systems.

00:07:02

Yeah, Jeeva, I think the spot on that it is, it is this is far more than just a a technical problem and I think that's you know probably why it is so complicated, Srini, you speak to so many CIO CDO's can I kind of ask the same question what you know from those conversations you're having.

00:07:24

Do you hear? Do you hear that message? Echoed or our? CIO's talking about something else.

00:07:27

Absolutely, Andy. I think while you know we are all talking about AI bias, you know the way give articulated it, a lot of the bias comes from data, right and you know, for all the factors that Jiva talked about in terms of the datasets, temporal bias etc, right and so apart from the data sort of fundamentally, you know data is created by humans. Most of the data, it's not auto generated. So the bias starts at that point in terms of capturing the data to, you know labeling.

00:08:01

As Jeeva said, to sampling to you know applying the you know data to the algorithms etc, right? So the recognition is coming in and it's likely gonna have a couple of implications. One is to make sure that there's human in the loop.

00:08:18

To, you know, understand and contextualize the data and the potential bias with it. The 2nd is also that the data collection mechanisms you know for the new sets of data that is being created.

00:08:31

You know it's going to change significantly where a lot of these contextual data points will need to start getting captured so that in future for the new datasets that are getting created, the bias is limited or at a minimum understood well enough so that the AI algorithms can handle the data better.

00:08:52

The other aspect also is that you know, look, most of us you know in technology when it comes to data, we focus on structured data which is data that's in our databases. You know ERP systems, CRM systems, etc.

00:09:09

But especially with the generative AI coming in and large language models, you know to create with other sets of data which you know we called as unstructured data.

00:09:20

Think about it. As you know your documents, your images, videos, you know audio so on and.

00:09:26

So forth now.

00:09:29

There's been a lot of effort in streamlining this structured data, but I think there's, you know, the CIO's and the executives are beginning to realize the importance of.

00:09:40

Handling the unstructured data as well, you know to understand the bias, explainability observability, etcetera. So this is the other priority that is beginning to emerge because you know if you know small things like you know the choice of words.

00:09:56

They differ right, you know, Jiva talked about temporal bias in the choices of words that were used, you know, 40-50 years ago is different from, you know, the accepted norm of certain choices of words. And that matters in terms of how algorithms function.

00:10:13

Images, especially if we are looking at the healthcare industry, you know there's a significant potential to apply AI for, you know, the good of the, you know, humans.

00:10:24

The images that are captured in emirs, you know 40-50 years ago, but the resolution is very poor.

00:10:31

So if you apply that to AI, the results may be sub-optimal. So these are just some examples in terms of you know how the CXO's are beginning to see the importance for applying similar concepts of data bias that we do on structured data to unstructured data as well.

00:10:50

Thanks really, Deva, just back to you again. That example of South Africa, I think it is so profound just kind of unpack that a bit more. Can you share any other examples of applications where or instances where data bias has had significant consequence?

00:11:11

So great question, Andy, for example. And I think the example that I pointed out to related to South Africa was more focusing on the tempo, the temporal bias that can extend, which is actually training the data set in specific to a time frame. But data bias. I think the example that you are asking for is. So for example that if you are we have gone through the COVID research and also right now most of the institutions have COVID impact population COVID impact data among the population, right. So if you were to train the models again.

00:11:52

Based on all the data that we have and if we train the model based on only one particular demography, leaving out the other demography. For example, we know.

00:12:03

Based on the data that we have collected people who belong to the African American community were mostly impacted compared to the other demography, whether it is still a, it is still a debatable data, but still at least the initial datasets have.

00:12:23

Actually proven that one particular demography has actually been impacted more widely than the other.

00:12:30

Right. So now if we are training the data sets excluding that particular one, imagine the impact it would have on people who are in entire America right now where these demography of people lives and so you will not be able to provide a better and also proactive care to these type of people, So that is why that anytime when we are actually putting a data by us, we need to make sure that we are including every single possible way that all the data related to the entire population data set where I keep using the word population dataset, it could because it's much more easier for people to get behind the line of thinking.

00:13:15

But any project that we are working on any AI model working on, even if it is for a missionary, we need to bring the entire both historical context all the way up to the real data port.

00:13:26

Thanks for that Jeeva. Jeeva that you mentioned the, you know there are there are lots of stakeholders involved in this. Can you just talk a little bit more about the role of those stakeholders, you know especially developers and end users in ensuring that we you know we maintain transparency in these models?

00:13:49

Absolutely. Andy. And I mean think about it for the last two years, this is one of the most talked about and most debated term since the AI models have actually started to enter into our enter into the industrial use cases, right. So of course AI governance is extremely critical.

00:14:10

For ensuring the transparency, compliance and also the trust in the development process right, the trust in the development and also the application of AI systems.

00:14:20

And in doing that, all of these people have a role to play both that not only the developers and end users, but also the policymakers and also the Academy and community as well to bring in all the feedback and to improvise this whole notion of AI transparency and also trust.

00:14:41

So from a developer standpoint, Andy, right? They are the people who are actually designing an implementation and I alluded to this earlier in the conversation itself that there are data biases can actually happen both by ignorance and also it can happen consciously because of the way that it is being designed as well. So I think it comes down to that very fact that.

00:15:05

How the developers are, you know, they are ultimately the people who are responsible for creating AI models and algorithms and they play a very, very critical.

00:15:14

In ensuring that the transparency and trust is actually maintained and the only way they can actually go about doing that is that any AI model that they develop should be explainable right at the end of the day, when you are pulled into. Inquire about a model, efficacy or otherwise, an issue. You, the developers, who actually designed and implemented, should be able to explain and interpret and Weld and document clearly about how they have actually trained the model and how they have developed the model and all that. And this level of explainable city. It rests solely on the. On the developer side of it.

00:16:00

Right. And equally, if you think about the end user side of it, one of the most important things that we have to remember that is that end users are the people who are finally using all these AI systems and they need to have active participation in this entire loop of feedback. They are the people who are testing it and they need to provide valuable feedback.

00:16:20

During this development pace so that all the insights that identify the issues related to the transparency, usability, fairness or actually communicated back to the purse and then they can actually fix it, right? So that is between the developer and the user. But apart from that, as both as a company, as an organization and also as a country, I think we need to have a formal framework related with policymakers and ethicists, right?

00:16:52

So we have seen the regulations which are being discussed, right? Now very widely and we have seen that policymakers are working to create a guide, regulations and guidelines and just like you know, when we web bolt actually came when Internet came about 20 years ago, we have seen that we have gone through this. We have undergone all these kind of policy making.

00:17:16

And also the governance and the regulations and everything else. Similarly, I think AI is in the early stages at this point in time and definitely the policymakers have a very good role to play in this one to making sure that the AI models are being developed and deployed.

00:17:35

Under a a good governance framework that provides transparency and trust to find fairness to everybody.

00:17:43

Yeah, thanks, Jeeva and the role of the framework, I think as we're developing this becomes really important in terms of being able to, you know, embed all of the learning and then, you know, socialize that out onto A to a wider audience really. Are you seeing any, you know, specific practical guidelines or or frameworks that are getting developed to help here?

00:18:09

Ohh absolutely Andy. So for example, if you think about it. To ensure that there is a responsible AI practice and to mitigate all the biases that we talked about, the fundamental very principled thing that very, very foremost thing that any organization can do is to establish the AI principles. Right. And this is.

00:18:31

Where the management team of that organization plays a very critical role in defining a clear set of AI principles that align with the organizational values. That's super important for that. And then now, once you actually established AI principle, then you can develop that AI governance framework right? It is to bring.

00:18:52

It is to connect the organizational values to the technology within a good framework operating model. I think both of these things will go into a that's the Super most important thing.

00:19:06

But apart from that, I think it's when it comes to the practicality of it. It is also investing in training, right, educating the employees, especially the developers and data scientists about responsibility. I and the consequences of missing something both ignorantly and also consciously and the impact it will have in the real world, all of these things should be, you know, educated. And I think every organization has that moral responsibility to invest in.

00:19:36

Training and then finally it is still. If you think about an organizational standpoint, I think there has to be a formal process where there is a formal feedback process where the questioners are actually sent to all the employees or whoever is the users of the AI system to.

00:19:57

Assets and to get feedback and assess all the potential biases.

00:20:01

That they have actually seen in the AI deployment so that they can actually identify and address all those kind of stuff. So absolutely from creating the AI principle to defining the framework to investing in training and investing in feedback process that brings the actual real world.

00:20:21

Experience back to the developers. All of these are going to be critical for ensuring that responsible AI and practices are actually being deployed and also how they can continue to mitigate the data by access handy.

00:20:36

Yeah, thanks. Thanks. Jeeva. Really just expanding on that. What are actionable steps you are recommending that organizations take to implement those frameworks and to mitigate the impacts of bias?

00:20:52

Thanks Andy, you know, we are all learning in this space. Right. I mean, this space is evolving. You know we are. You know I believe that we are still in early stages of you know realizing the true potential of the latest developments. So while this evolution happens, you know, I think we have regulations. On one hand, I mean European regulations, you know the executive orders in the US, each country is setting up, you know regulations as well. So and then you've got industry level regulations that are coming our best practices so to speak. The Academy, you know. The academic institutions are publishing a number of frameworks as well, so there are several of these, you know regulations which you have to comply with, best practices that you would want to leverage frameworks that you would want to apply into technology. You know they are there.

00:21:44

Now the step one is really about, you know, setting up a, you know, something like an AI council.

00:21:51

To start, you know, driving what it means for the organization.

00:21:57

In terms of driving AI responsibly, so Jeeva talked about it right needs to be embedded into the core values. So but you know there are all these regulations, best practices available. AI is going to be cross functional. It's not just about IT. It's going to be truly cross-function.

00:22:14

So, one step that you know we recommend we start seeing clients do a setting up an AI council and it is at a leadership level because a lot of this change needs to be driven with strong leadership. So then the second part of it is in terms of, you know setting up you know. Clear policies and guidelines. You know A to make sure compliance is no choice and B all to leverage the best practices and frameworks that exist. But you know they do need to be, you know, customized and tuned to a particular organize.

00:22:49

Right, so these definitions are extremely critical and this could be on various aspects, right from data collection to the technology choices to usage of technology to security of the data and the way the AI models are developed to how it is used. How it is explainable so? You know this, this team would really, you know, set up that for like a bit of words, you know, policies and guidelines, you know, for the enterprise. Then the next step is, you know, driving the adoption.

00:23:25

You know, it's not just about, you know, defining the policies and, you know, frameworks, but this is where you, you know, you would start looking at, you know, creating AI champions within the lines of business or within the IT teams that are driving the adoption of AI that are tightly linked. Up to the frameworks.

00:23:44

Policies and guidelines that are coming in and then finally a feedback loop because as we said, this basis evolving. So the feedback needs to go back into the, you know, into the Council to make sure that the guidelines, the frameworks you know and the policies are updated regularly.

00:24:02

You know, based on the real world evidence of how these policies, guidelines and frameworks are playing out and then continue to enhance there, it is also possible that some companies, you know enterprises over time, you know will connect it the you know, the A regulations into CSR because, you know being the responsibility to the society.

00:24:23

You know also would start meaning that you know we are applying AI responsibly into the businesses so this place is evolving, but these are some simple steps. There's a lot of hard work that goes into it, but a simple structured approach towards how enterprises can drive more responsibly. Yeah, on an ongoing basis.

00:24:44

Yeah, thanks. Trillion. I think from that conversation, there's two. There's two points that jump out for me. One is about Skilling and I think this goes well beyond.

00:24:56

And just a, you know, technical certification on the platforms to really understand, you know, data bias and ethics at a much, much more detailed level, I'd kind of love to see ethics and you know understanding data statistics on you know maybe it needs to be on every MBA. Every business course, every computing course, every degree there needs to be a grounding in some of these subjects.

00:25:26

And Srini, surely you also mentioned the human in the loop, yeah. Clearly we're at the early stage in this and having the human, not just the human, but I think the expert in the loop is key to make sure that we always question back the answers are and look at that explainability of the model and the transparency. Did you mention that very eloquently about just how important it was to be able to for the developer to be able to explain the model rather than just code it.

00:26:05

So I think this has been a you know, a fantastic conversation. Everybody, I hope you found this episode of The Elevate Podcast Useful. You've learned something. I certainly have. I think data bias, the bias in AI. These are serious challenges. And important as an industry, we really, really focus on them. I want to extend my thanks to Srini and Jeeva both for all of their thought leadership and partnership in this and for sharing all of those insights.

00:26:42

I'm not over. Discussing and exploring the potential of AI, so check back in again soon if you've got any questions or feedback, do feel free to reach out. So until next time, I'd like to thank you. Thank you for listening. Production team. Of course, you're always brilliant and Jeeva and Srini thank you very much.

00:27:01

Thank you, Andy.