Elevate Podcast series: Episode 4 - Moving beyond bias, making AI trustworthy

Welcome to elevate a podcast series by the HCL Microsoft ecosystem unit exploring the intersection of innovation and technology.

Andy says: Hello, welcome to another podcast in the intersection of innovation series. Innovation is all about three things, creating a safe, innovative culture, supporting collaboration across boundaries, and execution, creating value for customers doing good for society. Today, I want to talk a bit about AI ML, it's going to shape the future.

It's shaped it shaping our future today. It's providing new insights into operations into decision making processes. It's truly innovative. And development has been hugely collaborative. But is it bringing real value? Is it doing good for society? And that's something I want to look at today around how bias can bring in these unconscious impacts into the decision. Starting off with a couple of statistics that quite surprised me. Gartner predicts 85% of AI projects will deliver incorrect outcomes, that's most projects will have some level of error in in that outcome, whether that's data that's algorithms or down to the team. Lots of reasons for that second statistic 80% 78% of AI professionals have meant the huge amount of male experience coming into driving those models, and something that we should be really, aware of.

So, I'm really pleased today that I've been joined by Aruna Pattam - Head, AI & Data Science – Asia Pacific & Middle East, HCL Technologies and also joining us today is Poonam Sampat, Cloud Solution Architect, Microsoft. The question is not so much about can we trust AI? But more about what do we have to do to make sure that we can build trust in AI in ML? But before we go on Aruna? Could you just elaborate a little bit on the various forms of AI bias where they come from and some of the impacts?

Thanks, Andy, for the introduction, and it is a very important question that you just asked. But before going into the types of AI bias, I would like to just quickly touch upon what AI biases, AI learns by feeding huge volumes of data. And they look for patterns within this data. This helps them to make predictions, which are likely to influence the decision making, and impact those it is used.

Now, AI bias is the AI's version of discrimination towards a certain information, that impacts, a community, gender, or ethnicity. And now coming back to your question on the various sources of bias, bias can be introduced in many ways. And also, at various stages of the lifecycle. If the quality of the data that you use is not good, and it is influenced by human decisions, then even if that data has been used to build the algorithms, you're going to bring in bias, you can also introduce bias, by the way you're collecting and processing the data, such as over sampling or under sampling, where the data is skewed towards a particular group.

Humans, when they build algorithms can bring in bias, if they have an inherent bias, either consciously and unconsciously, then they're going to influence that into the algorithm that they're going to bring in. Also, we have seen, bias can be introduced even after the algorithm been released into the world with new data that's being fed into the live system. And if that data has got bias, the algorithm slowly learns from that data and start introducing bias. So, we need to be very careful on these types of bias. But most importantly, we need to be aware of the implications of bias as they have serious consequences to people and I can quote a few instances where we've seen AI bias in hiring, where Amazon has had to stop its AI hiring to because it found it was biased towards their female employees. And this is because that the data that was used to build the algorithm (it used the historical data about 10 years back) which was reflective of a male dominance in the tech industry, the AI algorithm started predicting more male applicant beings suitable, we have seen bias in policing, where we've seen an AI tool was producing our scoring between one and 10 to quantify if an individual will be rearrested is released.

And what they found is that the data that was used to build the algorithm was an issue. So, one of the features was a number of errors. And it was found that the male population was twice as likely to be arrested than that the white, and that, again, influenced the algorithm, it started predicting that more likely black people will be arrested. We have also seen a bias implication in the financial services industry, where a complaint was raised against Apple's credit card, and it was found that the AI system was predicting different credit card limits - based on gender, so one of the apple customers applied for a credit card, and he was given 10 times more credit limit than his partner, who had a much better credit rating than him. More importantly, we have also seen the adverse impact on health services, it was found millions of black people were impacted due to the racial bias in their health care algorithms that was used in the US hospitals. And it was less likely recommending black people from getting the immediate care, even though they had the same sickness and wanted immediate treatment. So, as I've seen, you know, given a few examples, where if you're not very careful, there can be serious consequences.

Andy says: Thank you for that. I think that's key that companies will have to take deliberate steps. It's, not something that can be just seen as part of the plan, I think we need to put a diversity and inclusivity track into everything that we're doing around AI, ML everything around data, it's just so fundamental to everything. So, does AI raise new challenges? Is this something that we've always had? But AI is kind of in the news, and everyone sees it? Or are there new challenges in fairness, inclusion, that that we need to understand?

Andy thanks, and it's a very good question diversity and inclusion is a very hot topic around the world, and AI and machine learning, I see no exceptions to it. As AI becomes more prevalent in our day to day lives, from voice recognition on smartphones to self-driving cars, it is becoming harder for developers not to consider bias, as they design algorithms that interact with people at work at home on the roads.

Yes, I do see challenges that AI can introduce to fairness and inclusion. But the AI community is already doing a lot of work. Working on the methods to help AI reflect all of humanity. We have seen some lot of changes here. We have AI self-check, which is designed to challenge itself and make sure its own thinking is not biased. We are also introducing humans in the loop so that the AI systems are developed and deployed with human oversight. There is also an AI fairness check that's been introduced where AI is being used to check bias in live data, even when it is streaming from sensors and other real time systems.

A big development I've seen is on the AI explain ability, where researchers are constantly working on ways where AI can communicate better with humans with more data points and evidence as to why a certain prediction has been done, and why a certain outcome has been recommended. We are also seeing AI systems being developed with diversity in mind and with a goal of increasing workforce diversity and making sure that all groups in society are part of the AI data and algorithms. What I've seen is that, up until now, developers have focused more on making AI faster, smarter, and more accurate. But now people are starting to realize and even the companies have started to realize that to make AI more powerful it should reflect all of humanity. That's how, you know there is a lot of new work needs to be done but It's really progressing towards that.

Thank you. And I love that idea of being able to go and ask the why question. Why did you do this, why we used to always challenging each other in a way we're used to this challenge between people. And that helps understand what the decision-making process is. It's a great way of learning. So, I think that the ability to go back to the system itself again, what was it in all that data and all those parameters and all of those weightings that got you to the decision? I think that's important to build that trust to be sort of, you know, a transparent decision-making process.

Poonam, I would love to hear from you on what you're doing in this area, how you're working with customers, to help them really understand the challenges and get to some solutions.

Thanks for that, Andy. And that's a very interesting question, because as Aruna said there is a challenge, and we need to address it. And I really think you know, right now, we are at a very extraordinary moment. And I think of it more like an inflection point, where the power and the reach of machine learning is rapidly expanding into many day-to-day activities, be it around healthcare, education, even while booking a cab, you know, when you're taking an Uber or a grab, everywhere, there is a machine learning algorithm working now.

These are exciting times. But there are also concerning problems like bias, stereotyping, and unfair determination. And this has been seen across a variety of machine learning problems around vision systems, about object recognition and NLP word embedding, I think Aruna highlighted the scenarios and implications of the same. So, these biases do matter. And I feel bias is the real AI danger. Now, at Microsoft, we recognize this, and we've been working in collaboration with researchers and practitioners around the world.

And we've come up with a responsible AI framework - right now, this is considered this as a guiding principle, you know, which we really follow internally as well, to build our products. For any of our AI products, we use these principles, which are around fairness, reliability, privacy, security, inclusiveness, transparency, and accountability. And what we've also done, we've taken this to the next level, and we've open sourced this responsible AI toolkit. Now this is to help the larger ecosystem, right now, what is this responsible AI toolkit, think of it as an interoperable framework for accelerating the development of responsible AI. Now, it integrates many tools in the areas of fairness, interpretability, error analysis, and causal decision making.

And even when you started, you spoke about how the error rates are there in these models. And that is what we want to eradicate and help people understand how to go about it. And this responsible AI toolkit, it's highly customizable, modular, interactive, and it can work with any machine learning code that is in Python. Now, I would like to call it like Lego pieces on how to build a responsible AI framework. Now, we have customers who have used this and been able to improve their models. This was a customer that we worked with very closely, and they used this responsible AI toolkit. And they were able to improve their loan decision application make it more fair, they were able to understand what the errors were, they used the fairness tool, which is part of this responsible AI toolkit to help identify and mitigate these issues. And with that thought, I would really urge every ML developer and architect to go and check out this so that they can make the models more inclusive.

Thank you. And I think that making that that model, open source is important. This is not the time if we're trying to build up trust in systems to have proprietary systems that are locked down. And I love the pillars. I think that's just what almost what decision-making processes should be like and it's kind of codified that, you know, for AI and ML but, but I think more generically, it's a great way to think through, you know, all of the decisions we make, should be fair should be reliable, they should have privacy and safety and inclusivity, kind of in, in the core of the decision-making process.

So, I love that. And, you know, so pleased to better work with Microsoft on, on these initiatives that can open sourced and, you know, back to my three sort of pillars of innovation, you know, do good for society. And so, so important. So that's brilliant. But before we close any final comments?

Sure, Andy, I think this is something that we all need to think about. Our systems will only work for everyone if they are designed with multiple voices in the room, every step of the way. So, it is very imperative that the entire process has these checks and balances in place. And that's what I would like to summarize, when I think about biases and how we can solve these biases in real world today.

Thanks, Aruna. Anything from you?

Andy, for me, the impact AI bias can have on our lives and businesses are very significant. So, we must be aware of the different types of bias and take steps to prevent them from influencing our decisions. Only then can we reap the benefits that AI has to offer.

I think it really, I think you're spot on awareness is the key to this and we need to have everybody's voice. You know, as well as the data scientists and developers, the society that those decisions are impacted. If through that framework, they can have a role in playing and shaping the decisions. I think that is what starts to build trust in systems. And the fact that we can go back and ask why, again. I'm very excited. I think this is an area where we will be able to do good for society. But at the same time, there's a lot of risk there. Thank you.

Thank you for listening to elevate a podcast series on the intersection of innovation and technology by the HCL Microsoft ecosystem unit. We will be back with our next episode soon.