WHOSE BIAS IS IT ANYWAY?

In September Pegasystems and UCL ran a virtual roundtable on the topic "Do AI biases and human biases overlap more than we think?", presented by Peter van der Putten (an assistant professor of AI at Leiden University and director at Pegasystems) and Dr Lasana Harris (Senior Lecturer in Social Cognition at UCL).

It's interesting to hear from the experts about both the potential and the limitations that AI tools bring. We've seen that AI tools can perpetuate biases that exist in society, but is that any less true of humans? Do we, and should we, expect more from computers than we do from people?

AI and customers

AI, in the sense of machine learning approaches to automation of particular tasks, is increasingly a necessity. The pandemic is a good example of a crisis which demands systems, like track and trace, which are able to handle vast amounts of data efficiently.

When dealing with customers, Ai systems often fail because they are not able to feel and express empathy in the way that humans are. As Peter van der Putten commented,

"People view companies as if it is a single organism or as a 'person'. This requires not just intelligence but also empathy. Sense my emotions in the moment, learn from interactions to understand my needs, but more fundamentally, put yourself as the company in the shoes of the consumers: say if we want to put some message or nudge in front of a customer, promote what's right for the customer not just what's right for the company."

Customers are, for the most part, sceptical about organisations' willingness to do the right thing for customers, beyond what they're legally required to, and that creates opportunities for companies who can show customers that they do.

Pega shared the example of CommBank, who introduced a Customer Engagement Engine, using AI to proactively select the "Next Best Conversation" that is best suited to the needs of each customer at that time and through that channel. As Peter explained,

"...the library contains a wide variety of messages, in line with the mission. Not just selling the product of the month and personalised sales recommendations; also warning about credit card points that expire, how customers could avoid upcoming fees and

Stephen Hampshire
Client Manager
TLF Research
stephenhampshire@leadershipfactor.com

charges. But also beyond their products and services: the Benefits Finder identifies specific government benefits specific customers qualify for; or emergency assistance when customers live in an area affected by bushfires. Since COVID-19 hit, it communicated 250m COVID-related messages to customers, from payment holidays to home loan redraws."

> "The most important thing is that when companies use AI, they must balance their self-interests with those of the consumer."

Brands as people

Why do customers think of organisations as if they are people? Because they spend millions of pounds on advertising to position themselves in that way. One of the biggest causes of customer dissatisfaction is the disconnect between the friendly, personal, brand they're promised in the adverts and the impersonal treatment they often receive in practice. Dr Harris commented,

"The most important thing is that when companies use AI, they must balance their self-interests with those of the consumer. When deciding how they want to use AI they need to consider whether it will impact their brand and their reputation."

"People typically aren't very trustful of their information with companies and AI presumably can help smooth some of that transition if used appropriately. I don't think that perception of AI in general is that it's evil. I think that comes

out when it is used in ways that sort of threaten things that people have, like their privacy, for instance. So it's really about the goal of the company."

What should the future of AI be?

If customers are increasingly sceptical of the benefit to them of AI tools, and as wider society begins to worry about the impact of algorithmic bias, now is a good time for organisations to consider how best to deploy AI solutions in a way that is good for customers, and society, as well as their bottom line.

As Peter commented, actions are more important than words here:

"Just defining AI principles is not enough. I think there are two things which are really important. One, you need to translate these principles into something tangible. When you say you need to be transparent around automated decisions, you need to offer some form of automated explanations on how this decision was reached. If you say we're against bias in models, but also in automated decision in systems, you need to have an ability to measure how much bias there is in those decisions in the first place."

It's also about making sure that AI tools are used in the right way, and in the right places. Used well these approaches have the potential to deliver much quicker, more responsive, more personalised customer experiences at scale. Customers will make up their minds based on the results they see. The aim, as Lasana says, should be to…

"Improve the life of your customer somehow, and the AI can facilitate that…given the power and the influence of AI, AI can make decisions across thousands of customers very quickly."

Where does the bias come from?

Algorithmic bias is not inevitable, but something which comes about because of the way we build and train AI models. Those biases reflect tendencies in the data, in other words they may recapitulate systematic biases in society, but they may also be exacerbated by who works in tech and the way they think.

So what does cause these biases, and is it

something that we can easily take steps to prevent? Peter highlights three fundamental, related, problems:

- *"In the same way that humans see bias in society which then reinforces our own biases, it is the same for AI.*
- *Through the data that we use to train models, through bias in decision logic driving automated decisions, through more mundane problems such as data issues.*
- *That's not an excuse to blame it on the data – the more systemic issue is not having an eye open for the bias that could occur or not having the tools to detect and fix it."*

> "I don't think that perception of AI in general is that it's evil. I think that comes out when it is used in ways that sort of threaten things that people have, like their privacy, for instance."

This is really important, and links into the points made in Artificial Unintelligence (our book review on page 31). To see algorithmic bias as merely a problem that reflects the training data, and therefore as society's problem rather than AI's problem, is missing the point. Peter continues, with some examples:

"The more systemic underlying issue is that ultimately it's humans that build AI systems. So

the systemic problems are added that people are maybe not aware enough of, or bias problems happen, or the systemic problem could be that people don't care enough."

"One example is the 2020 A-Level results (in this case, not AI, but algorithmic bias). Boris Johnson blamed a "mutant algorithm" for the A-level and GCSE grades. You can't just blame it on the algorithm. Algorithms are not silver bullets, nor are they inherently evil. And algorithms are certainly not objective, nor 'back boxes' we can shift any blame to."

"Another is a study in Science in 2019 which reported on a predictive model used across the US to identify patients for preventive care and care management programs, clearly an example where AI was used with the best intentions. The problem is that the model predicts future healthcare costs, and in the historical data used to build the model, considerably less money is spent on black patients that have the same health conditions as white patients. By correcting for the bias in the healthcare data set, more than two and a half as many black patients would be eligible for a care management

"If you say we're against bias in models, but also in automated decision in systems, you need to have an ability to measure how much bias there is in those decisions in the first place."

program. Bias was caused by how the data set was defined for modelling."

What can we do to prevent bias?

To build algorithms that are unbiased requires active work, and making sure you understand the nature of the data that you're using. Knowing how the data was collected, and the nature of the society in which it was collected, is as important as being able to build an efficient algorithm. As Lasana comments,

"I think when discussing bias, it's really important to understand that the bias exists all around us...If there's no bias detection mechanism and there's no person who's aware of these biases intentionally looking to see that they are not present in the AI, then the AI is going to appear to be biased."

"In reality, the way to combat social bias is to be aware of your own biases – the same thing is true for AI. Therefore, those who are creating AI need to be aware of their own prejudices."

The conclusion is clear: if you want to build AI algorithms which are free from bias, then you're going to need to build transparency and bias detection into your systems. This can't be done passively, but needs to be consciously approached with an understanding of the potential biases that training data may reflect.

You also need to evaluate the decisions or predictions that your algorithms are making, and make sure that they are fair.

Are we being unfair on the machines?

People within tech often feel that all this is a bit unfair, after all machines are, by definition, free from bias themselves. If a computer learns to replicate decisions which are biased, based on a stack of data about how humans have made decisions in the past, then that's hardly the computer's fault, is it? And yet we seem to be suggesting that computer decisions should be more heavily scrutinised than their human colleagues. Is that right?

To some extent that's our pro-human bias speaking, but there is also the

"Humans are sometimes eager to push responsibility to an AI algorithm, which is not correct."

question of scale. A biased human makes far fewer decisions than a biased machine. Nonetheless, it's important to remember that, however flawed a computer's decisions, the fault remains with humans, as Lasana comments:

"Humans are sometimes eager to push responsibility to an AI algorithm, which is not correct. AI algorithms are built and trained by humans, based on a range of choices made by humans."

Perhaps the most important thing of all for a customer, whether the decision was made by a human or a machine, is that it seems fair and is explained. As Peter says,

"For a customer being declined for a loan it doesn't matter that much who made the decision, the human or the AI. She or he wants the loan and didn't get it, so wants to get an explanation and wants that decision to be fair."

"Make the customers feel that for every single customer and every single interaction you're really trying to do the right thing for them."

Ultimately, like everything else in the customer experience, what really matters is that customers believe that you are on their side, and have their interests at heart. If AI is serving that end, then it has the potential to contribute to excellent new customer experiences, but it can't do that until we take a clear-eyed look at the biases that we're building into decisions and predictions. To do that, we're first going to have to face up to our own biases.