**Financial Inclusion TechSprint (May 2024) video Transcript.**

**Team 10 – Zuhlke**

**Delegate 1**

Hey. Hi, we are Zuhlke. We are an innovation and transformation partner. We have almost 2000 employees globally and we have a proven track record in the finance, government and health markets.

Our solution is an innovative way to understand and address bias. Those who are deploying AI technology need to understand that bias can exist and be introduced at every stage, and we want a world where firms can understand where their bias negatively impacts consumer groups and can make changes that lead to good outcomes. Our solution provides transparency where bias exists across the end to end workflow and supports the choice of AI solutions.

So we present to you the Bias framework, which allows you to investigate the trade off between usability and bias to find the most appropriate techniques for a given situation.

**Delegate 2**

The truth about developing decisioning algorithms is that it's a complex pipeline that thread together the data that you input to them, your choice of models to survey to select which one or your choice of decisioning algorithm, and then the metrics that you choose to measure the quality of the outcomes. Not only then do you decide what to deploy, the huge complication is that bias can enter that pipeline in any part of that pipeline and many parts of the pipeline, and it gets convoluted as it goes through the pipeline to the outcome. It's a really complex situation.

What we do with our framework is we take holistic view of the problem and at a high level what we're doing is jiggling around aspects of each of the steps to understand its impact on the output metrics. And we do that by making visualisations so that we can judge where the largest biases are coming from and then iterate so that we can address that.

**Delegate 1**

So we'll provide you with an example use case based off the sandbox data available to us to bring our solution to life. Imagine a credit rating agency determined an individual determines an individual's credit score using an AI model.

Firstly, the agency can configure the framework. They can decide which version of their data set they want to run the data the framework on decide which variables are to be used for training, protected for debiasing and which is the target. They can select versions of models and determine how people are determined to be in the favourable or unfavourable class. Lastly, they can choose from our suite of debiasing techniques to run with their models. Each of these debiasing techniques will alter the predictive outcomes of the model and it is that impact that our metrics measure.

These will then see a centralised view of relevant metrics. They'll get a variety of both accuracy and bias metrics so they can look for consistent and relevant performance in the debiasing techniques they've run. They may will also want to see the real world impacts of these debiasing techniques on their data. So for example, mapping the change in credit score ratings across the across the UK for the debiasing technique Prejudice Remover.

We'll provide a variety of visualisations to give the user a really holistic view of the debiasing techniques. And the next steps is for the user to iterate on this process, make known alterations until they come up with a solution they're really happy with.

So what have we done these past three months? How have we implemented this with the Sandbox? We began by creating reproducible pipelines that output multiple datasets. And this is really key because you need to be able to measure the accuracy and bias of each of the datasets you've outputted. We use the Women's Economic Empowerment Tech Sprint and automated push payment data from the Sandbox. And we introduced the Indices of Multiple Deprivation from the UK when it came to joining these data sets together. And we identified 3 different ways you could join them by sort of characteristics, geographically or with the IMD. We didn't select just one of these because to choose one of these datasets would be to introduce our own bias into the pipeline. This is why reproducible pipelines are so important. They allow you to produce multiple datasets so you can measure the accuracy and bias of each one to find your optimum.

**Delegate 1**

So in the sandbox, what we've done is we've covered it up an initial wide range of bias related metrics and accuracy related metrics. And what we're going to zoom, we're going to do is zoom into one of these to look at the what we've been doing.

So for one particular bias metric and one particular accuracy metric, we'll just choose one model and plot the results of those metrics there. And the error bands just represent how uncertain we are about where to place the points. Now with that one model, what we do is induce systematic changes that make it less biassed. And what happens when you do that is the accuracy goes down.

This is the bias accuracy trade off. There's this underlying tension in this problem of training algorithms. Now let's have a look at what happens when we add in a view of 1 particular debiasing technique. This output is quite a quite a positive one in that the accuracy has not been reduced, but it has become less biased. So it's a candidate for quite a good algorithm to use. But there are lots of bias mitigation techniques, so let's plot some more and see what happens.

On the lower right. There's an example of one that's not so good because its accuracy is reduced, but it hasn't got less biased in the upper middle of the plot with the green. Overall, there are some results there which have got quite high accuracy and some debiasing. Notice the importance of the error bars in this problem, because you would not want to distinguish between the results of that green Oval because the uncertainties are larger than how different the points are. So that's the crucial role of the error bars.

Now, all of that was run on one particular data set configuration in the sandbox. And as Tabby was just describing, there are various ways of doing that. And to your point in the panel, the focus on the data is so important. So watch what happens if we just change the data set configuration.

All of those results move rather along. That just illustrates how sensitive the problem is to the input data. So how are we going to implement this? So the points about collaboration, it is actually a collaborative effort with this tool to get the right people around the table to be transparent about how certain decisioning algorithms are arrived at. So it's not just technical people like ourselves who are running the code, it's sitting with you as the organisation wanting to implement the decision making process. It's transparency to the regulator and all of the other relevant stakeholders in the problem as well.

So the benefits of the Biassed framework are that it's a holistic view of bias in the pipeline, it shows how bias mitigation impacts customers, and it's transparent, auditable and robust decision making.

So taking it forward, what are our next steps?

Well, the key thing is that we would propose a an offering to prospective clients where this is a service where we work with you to arrive at that transparent decision making. One of the key things about the implementation process is that fairness and bias is a mentality from the outset of looking into these algorithms. It's something that goes all the way through the technical process.

I'll leave you with a quote from a local historical figure, Lord Kelvin. Uh, we were, we didn't know what you're going to say earlier on about measuring, but to measure us to know, and that's one of the fundamental underlying principles in this AI development process is that to measure every step of the way and understand the systematic shifts is a critical aspect of the transparency of decision algorithms.

Thank you very much.