Transcript

Title: Webinar: Social Data in Action - We are AI: Towards Robust Public Participation in ADM Oversight
Creator: Prof Jane Farmer, A/Prof Anthony McCosker, Julia Stoyanovich
Year: 2021

Audio/video for this transcript available from: http://commons.swinburne.edu.au

JANE FARMER: So hi, everyone. Thanks so much for being here. This is our fourth webinar in this series around social data in action. And today we are very pleased to welcome Julia Stoyanovich from New York University, who's currently on holiday somewhere in rural Pennsylvania. Thanks for being here.

Julia is going to talk about We are AI Towards Robust Public Participation in ADM Oversight. As usual, this webinar is brought to you by the ARC Centre of Excellence in Automated Decision-making in Society and the Social Innovation Research Institute of Swinburne. So I might just introduce Julia briefly and then go on to the mechanics of the webinar, and then we'll hand over to Julia.

So Julia Stoyanovich is an assistant professor of computer science and engineering and of data science at New York University, although interestingly, we've just been having a great discussion about art and how important that is. And we'll hear about that today. She directs the Centre for Responsible AI at New York University, a hub for interdisciplinary research, public education, and advocacy that aims to make responsible AI synonymous with AI.

Julia's research focuses on responsible data management and analysis. She established the data responsibly consortium and served on the New York City Automated Decision Systems Task Force by appointment from Mayor deBlasio. And should we move on to the next slide, Paul?

So just like to do an acknowledgment, acknowledge that I'm hosting this webinar from the lands of the Wurundjeri people of the Kulin nation, and I acknowledge the traditional custodians of the various lands of which we are all working in various parts of the world and Australia today and the Aboriginal and Torres Strait Islander people participating in this webinar. I pay my respects to elders past, present, and emerging and celebrate the diversity of Aboriginal peoples and their own good cultures and connections to the lands and waters.

And so while Julia is speaking, we really encourage you to come up with questions and stick them in the chat as we're going along. And at the end, Anthony will deal with these questions, mediate your questions, and you might get to ask the question yourself. We really do encourage you to ask these questions. This is where all the great stuff comes out, so please do.

And if you need to, please do raise your hand during the Q&A session if you want to add to it. And we are recording this session. The previous sessions are available on YouTube. And if you have any challenges with this being recorded, please get in touch with Paul Lavey the email address given there. So now handing over to you, Julia. Thank you so much for being here. The floor is yours.

JULIA STOYANOVICH: Thank you very much, Jane. It's a pleasure to be here, and I actually was thinking that depending on how quickly I get through the first part of the presentation that I might do a round of Q&A in the middle if that's OK with everybody. But let's see how things progress. So please confirm that you are seeing my slides and not the notes. Yes?

AUDIENCE: All good.

AUDIENCE: Yup.

JULIA STOYANOVICH: So once again, a pleasure to be here. I'm Julia Stoyanovich, and today I will discuss some of the recent public education and public engagement work at the Centre for Responsible AI at NYU that I direct. Just a couple of words about the Centre, although Jane already gave an intro to us, our goal is to build a future in which responsible AI is synonymous with AI. And our work is focused on applied interdisciplinary research, policy, and public education and public engagement.

Here I wanted to acknowledge the team of the Centre for Responsible AI. I co-founded the Centre together with Steve Kuyan, who is also director of entrepreneurship and managing director of the NYU Tandon Feature Labs. This is a startup incubator that has been in existence for several years at NYU at the School of Engineering, and it has had tremendous impact on the startup ecosystem in the city.

We also are fortunate to have Eric Corbett and Mona Sloane, and they appear here in these pink frames because they were instrumental to much of the work that I will present today. Eric Corbett is a postdoctorate research fellow, and he specialises in computer science and in particular in human computer interaction. Mona Sloane is a social scientist, and she works very closely with Steve and me on building out the agenda for the Centre, and her research interests are participatory and collaborative design.

Also on our team are Raoni, who is a postdoctoral researcher with a background in data management and computer science, Eleni Manis, she's a political philosopher. Falaah Arif Khan is our artist in residence, and she is a machine learning expert and also, as you will see, an artist. And the images that I am using in my presentation today are all due to her. And Maria Grillo is a person without whom none of the public engagement activities of the Centre would have been possible.

We also have several very talented graduate students on our team whom I'm listing here. So let me, before diving into the main topic of the presentation, tell you a little bit about how I got to be interested in public education and public engagement, specifically around automated decision-making. So my research is I'm a computer scientist and a data scientist, and my whole community is data and knowledge management or databases within computer science.

So this is one of these core computer science communities that has existed and has been successful from the 1970s. And what we do is we build systems. And my own research is along sort of three threads. One of them I'm highlighting here because it's most relevant to the topic of today's conversation, and that is responsible data science, or RDS. I'll say another couple of words about this on the next slide.

I also work on combining data management with computational social choice, where we reason about elections and winners in elections under incomplete information. And the final line of work that I'm engaged in is kind of the most big data and systems-oriented line of work, and there we ask a question about how we should represent large evolving graphs and ask questions about their evolution.

So what is responsible data science? This is the most relevant topic of my research due to the talk today. Here the work that my colleagues, my students, and I do is under this label of data equity systems. And here we are inspired by a seminal paper from 1996 by Batya Friedman and Helen Nissenbaum, who identified three types of bias that can arise in computer systems. And these are represented here as a three-headed dragon.

Pre-existing, technical, and emergent is what the heads read. Pre-existing bias exists independently of an algorithm, and it has its origins in society. Technical bias may be introduced by the operation of the technical system itself, and it may exacerbate pre-existing bias. And finally, emergent bias arises in the context of use of a technical system. And it may be present if a system was designed with different users in mind or when social concepts shift over time.

In my work, I aspire to fight the bias dragon with the help of data equity, represented here as this female knight with three swords. So what do I mean by data equity? Well, first, equity as a social concept is about treating people differently depending on their endowments and needs to provide equality of outcome rather than equality of treatment.

This concept, equity, lends a unifying vision for much ongoing worked operationalized as ethical considerations across technology, law, and society. My colleagues, Bill Howe, H.V. Jagadish, and I overlay this concept over data intensive systems, and we get data equity. So what is data equity? We see it as having three major facets, one for each sword of this knight.

The first is representation equity. It refers to deviations between the data record and the world the data is meant to represent, often with respect to historically disadvantaged groups. The second is access equity. It's concerned with having access to information, including features, data, and models that are needed to evaluate and mitigate inequity. Finally, outcome equity refers to downstream unanticipated consequences outside the direct control of the system. And in the work that I will discuss today that is, once again, around public engagement and public education, outcome equity specifically comes into focus.

So in addition to the technical research, which I want to discuss today beyond what I already said, I've also been developing and teaching courses on a new subject called responsible data science at the Centre for Data Science at NYU, New York University. These courses are in their third year right now, and they are being offered to both undergraduate, as of 2021, and graduate students. These

are technical students. They study computer science or data science as their major or minor concentrations.

And as you can see from the photo on the right, this course has been able to attract a very engaged and also very demographically diverse group of students. All course materials are publicly available on GitHub. I encourage all of you to reuse them and to let me know what you think.

So I briefly mentioned that the work I'll discuss today is a natural consequence of my academic research and of my educational activities in the university setting. But perhaps an even more important reason that I am now involved in public education and public engagement is my almost accidental involvement in efforts to regulate the use of automated decision systems, or ADS, as we call them here, in New York City. So here's, briefly, a story on that.

New York City, where I live, prides itself on being a trendsetter in many things, including architecture, fashion, the performing arts, and as of late, in its very publicly made commitment to opening the black box of the government's, city government's, in this case, use of technology. The city was the first government entity in the United States to attempt this and, as we will see immediately, with mixed results. In August 2017, Council member Vacca of the Bronx proposed a local law that would compel all New York City agencies that use, and I quote, "an algorithm or any other method of automated processing system of data," to post the code publicly online and to allow users to submit their own data and see what the agency's code would return.

So this was a very radical proposal, a bold move, with lots of people and organisations making lots of passionate statements for and against it. And I was one of these people you can see a photo of me here on the right, together with Julia Powles on the steps of New York City City Hall. We are going in to testify on this proposed bill.

And Julia wrote a wonderful article about this bill in the New Yorker. I'm showing it here on the left. And I'm also including at the bottom a link to my own testimony. So fast forwarding now to January 2018, New York City passed a law-- it's called local law 49 of 2018-- in relation to automated decision systems used by agencies. And this law was based, but very loosely based, on Vacca's original proposal that I showed you.

I'm summarising the law on this slide. They define an automated decision system, or ADS, as "a computerised implementation of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions." The law mandates that the task force be put in place that surveys the current use ADS aides in city agencies and develops procedures for the following three things.

The first is requesting and receiving an explanation of an algorithmic decision affecting an individual. So this is an individual who is being affected by a decision, like they applied for their child to go to a particular school, and then the decision that was made was to give them a spot in another school. And they want to understand why the decision was made, and an explanation should be furnished to them.

The second is interrogating ADS for bias and discrimination against members of legally protected groups. And the third is allowing the public at large to assess how ADS function and are used by the

city government and also for archiving ADS together with the data they use. So once again, this task force wasn't going to actually provide mechanisms for giving this type of disclosure and for conducting this type of interrogation, but rather they were going to give recommendations based on which procedures would be developed.

So the passing of this law and the appointment of the task force that is illustrated on the left was celebrated by the mayor. And there were lots of expectations on the types of public disclosure about ADS that will take place in New York City as a result. The task force was comprised of New York City government agency representatives and also of external people, academics and representatives of various stakeholder groups.

And to my surprise, and also delight, I was appointed to the task force. And you can see me here in this image, and you can also see several other people who you may know. Next to me is Solon Barocas. He's a prominent researcher into space, of course, and a dear friend.

You can also see Jeannette Wing. She is a very well-known researcher in computer security. Maya Wiley is one of the people here. She is one of the mayoral candidates right now in New York City. And Vincent Southerland is a law professor at NYU.

So I already mentioned that the task force, of course, the results were mixed. And so halfway through the work of the task force, about halfway-- and the task force was appointed for 18 months-- several of us, including Solon Barocas and me, saw the need to speak to New York City government through the press to demand that information be disclosed so that the task force can do its work. So here I am giving a representative example from an article that appeared in The Verge that quotes me as saying that if no examples of actual ADS that the city uses are forthcoming, then there was really no point in forming the task force at all.

So essentially, what was happening-- and there is also a link to a testimony that Solon Barocas and I filed jointly-- was that the task force turned into more or less a formality, a formal exercise where the city was saying that they are ready to start disclosing so that we can come up with frameworks, but it didn't seem like they felt comfortable actually giving information to us to give us the necessary details. Be it as it may, now fast forward into November 2019, this is the end of the work of the task force, we did deliver a report that was then made publicly available. And at the same time, our mayor, deBlasio, issued an executive order that called for establishing an Algorithms Management and Policy Officer, or AMPO, in the city.

And the idea there was that the task force's report essentially ended up being very high level and that this AMPO's office would take wherever the task force left off and would come up with more concrete, more actionable recommendations, frameworks for city agencies to follow. So to summarise the report, we outlined several principles. One is that we should be using ADS only where they promote innovation and efficiency in service delivery, not because they are available to us. There actually needs to be a case made for an ADS to be used for it to be procured and deployed.

Promoting fairness, equity, accountability, and transparency in the use of ADS and reducing potential harm across the lifespan of these systems. And then we also came up with recommendations, and these were to formalise ADS management functions, build the city's ADS

management capacity, and broaden public conversation on ADS. And this third recommendation came with additional details about there being a need to invest in public education and public engagement so that indeed members of the public would have an opportunity to speak and to see what kinds of disclosure would be useful to them. And we will get back to that point in just a couple of minutes.

So stepping back for just a second, should we even attempt to regulate automated decision systems, ADS? For example, if we decide to regulate, should we use the precautionary principles, that is better safe than sorry? Or perhaps we might adopt a more agile risk-based method like algorithmic impact assessment.

And then whatever general framework we agree on, how do we go from general principles and recommendations to something that is concrete and actionable? One of the ways in which I remember the New York City ADS task force meetings is that everyone had the line that they would insist on and repeat at almost every meeting. And I was one of these people. I, of course, also had a line, and my line was, we need examples.

And I truly believe as a data scientist and as a person that we cannot figure out how to do things without concrete and difficult examples. So let's dive right in. And I'm going to discuss an example here. My example is going to be automated hiring systems.

So what are these systems? Since the 1990s and increasingly so in the last decade, commercial tools are being used by companies large and small to assist in the hiring process. This process is depicted on the right of my slide, and it has been aptly described by Miranda Bogen and Aaron Rickett as a funnel. It's a sequence of data-driven, algorithm-assisted steps in which a series of decisions culminates in job offers to some individuals and rejections to others.

And as stated by Jenny Yang-- she is the former commissioner of the US Equal Employment Opportunity Commission-- this is a federal agency that oversees these practises-- automated hiring systems act as modern gatekeepers to economic opportunity. So they are extremely important, of course. Why are these hiring tools so attractive? Well, first, they help employers hire more efficient, source and screen candidates with less paperwork and faster, and presumably select candidates who are likely to do well on the job.

And these tools are also meant to improve efficiency for job seekers, allowing them to apply with the click of a button and matching them with relevant positions and also facilitating the interview process. Another thing here is that we hear this argument time and again that the use of these tools, of automated hiring systems, can help us improve workforce diversity. And indeed this argument is being made very frequently, and that is because humans are biassed, we have no choice but to use machines to step in and hire on our behalf.

I am showing you here an excerpt from a famous study by Marianne Bertrand and Sendhil Mullainathan, who, back in the analogue days of 2004, manipulated perceived gender and race on resumes and observed that applicants who were perceived as male and white received substantially more callbacks on the very same resumes. So the question here is, are algorithmic hiring tools living

up to their promise? And of course, we have seen many indications of things going wrong. I will not dwell on this here in the interest of time.

Many of them I'm sure are familiar to you, where it was shown that, for example, Amazon's resume screening AI was prioritising male candidates and downgrading female candidates, also that women are less likely to be shown ads for high-paying jobs. This is on the top left in The Guardian. And there's also evidence that these tools discriminate against individuals suffering from disabilities even if they are well-qualified for the jobs.

And finally, some of these tools that are used as part of the informal background check process show very strong racial bias. So this is on the top right, a study by Latanya Sweeney, who showed that when you Google with African-American sounding names, you get ads that are indicative of criminal records, even when you control for whether the individual in fact has a criminal record. So not dwelling on this here, once again, examples abound on illegal discrimination, unlawful discrimination in which employers would engage with the help of these tools.

So in addition, or perhaps even before worrying about bias and discrimination, we should also ask whether these tools work in fact. Are they picking up useful signal from the data, or are they an elaborate coin flip at best? And here I'm depicting Arvind Narayanan. He is a computer science professor at Princeton, and he gives this brilliant talk about AI snake oil, using algorithmic hiring tools as one of the prominent examples.

So if a tool helps improve workforce diversity, it admits a sufficient number of candidates of each required demographic or socioeconomic group but its decisions are otherwise arbitrary, can such a tool be considered fair? And what if these decisions are worse than arbitrary? What if the tool is picking up signal like a person's disability status and we have no way to know that this is happening?

What if it's measuring something about the job applicants that we have no reason to believe to be relevant for the job for which they are applying? These are all important questions for us to be asking. And so following up on that, I'd like to bring to your attention another attempt, again in New York City, but this time specifically to regulate the use of automated employment decision tools, as they are called here. If this law passes, it will apply to the use of these tools, both in government and in industry, and it will prohibit the sale of such tools if they were not the subject of an audit for bias.

The law will also require any person who uses automated employment assessment tools for hiring and other purposes to disclose to candidates when such tools are used to assess their candidacy and also to tell them what job qualifications or characteristics the tool used to screen. So this bill, if it passes, would give us unprecedented insight into the use of these tools and the reasons that they are used and the types of reasons that they pick out about candidates. And this bill in particular, and the use and regulation of automated hiring tools more generally, has been the focus of the public education and engagement work of the Centre for Responsible AI at NYU.

And this is why I spent quite a bit of time discussing it here. So if you'd like to hear more about what I, in particular, think about this proposed bill, please take a look at the New York Times opinion piece that I wrote together with Alexandra Reeve Givens and Hilke Schellmann. I'm highlighting some points here about public disclosure about the need to consider the circumstances of people who are

multiply marginalised, the need to ensure that the tools actually measure what they claim to measure, and that these characteristics being measured are relevant for the job. And the final point that we make is that this information needs to be available to people before they apply, not only after they've applied.

So with all of this as sort of lengthy background, let's now dive into some of R/AI's public engagement activities on hiring ADS. We ran a series of such activities, and I'm highlighting some of them here. One of these is a session at the Queens Public Library's Jobs and Business Academy. There's another session that we ran as part of the New York City Open Data Week, and then I'm also highlighting here a very well-attended webinar that we ran that was conducted by Schneps Media.

All these events started with us setting the stage, in much the same way as I just did today, about automated decision systems in general, about algorithmic hiring tools. And also, we gave participants information about what algorithms are. What is data? What are algorithmic decisions? And I will talk about that in just a few minutes.

And then after setting the stage, we would respond to questions and ask participants about their concerns. And with the help of these events, we were able to reach a pretty broad cross-section of individuals seeking employment or just those being interested in algorithmic hiring generally in New York City, ranging from folks who attend the Jobs and Business Academy in a housing project in Queens to those who attend New York City's Open Data Week. So these are typically either citizen scientists, citizen hackers, or technology professionals.

So having run these public engagement activities, we summarised our learnings in a public engagement show reel. And this show reel is available, this document, at the link that I'm including below. And here are some of the most prominent public concerns that we heard, and we include them in the reports as quotes. Highlights here are mine.

One person said, "I would like to know how to minimise bias when I apply as a racial minority." Another said, "Is there any way to tell if a company or job posting is using AI?" So both of these points speaks directly to what this law 894 would require if passed. The third person says, "The employer should be responsible for ensuring that tools are built and used appropriately."

And another quote here is, "I don't think I have a choice not to share my data." And this is one of the comments that we heard multiple times, that employers choose to use these tools while employees are subjected to them. They don't really have a choice as to whether or not they are evaluated by a machine, and they don't even know whether they were evaluated by a machine in fact.

Here are a few takeaways that we summarised based on these activities. The first is simply that New Yorkers are interested to learn about the risks of automated decision making. And the second is that New Yorkers want to be engaged in the design and use of hiring and employment systems.

As one of the points that we summarised here that pertains less to the public engagement activities and more to the testimonies that were delivered and heard at the hearing for this bill, here we say that experts support this bill, but also outline key areas for improvement to make the bill effective. And you can see the details of the show reel at the link below. So where do we go from here?

We are reminded about the importance of what I termed outcome equity earlier. Outcome equity is concerned with whether and how we can assist and mitigate inequities that are outside the system's direct control. We can support outcome equity, I claim, only if we are able to hear from and work together with individuals who are impacted by automated decisions to empower them to change the way that these decisions are made, to change the way that these systems are designed.

So this is my outcome equity sword as part of the beta equity cartoon. So my thinking about outcome equity and about public disclosure centres around this metaphor of a nutritional label. The data management community, which is my intellectual home within computer science, has been studying systems and standards for metadata and provenance and transparency and explainability for decades. Public disclosure requirements are once again bringing these favourite topics into the spotlight, but with a twist.

The twist here is that we are now compelled to develop disclosure mechanisms not only for domain experts, not only for technical individuals who are building these systems, but perhaps most importantly for members of the public. To differentiate the nutritional label from more general forms of metadata, my colleagues and I articulate several properties. We think that labels should be comprehensible. They should not be a complete and therefore overwhelming history of every processing step that is applied to produce the result and decision in this case.

The information on a nutritional label must be short, simple, and clear. Labels should be consultative, providing actionable information, not just descriptive metadata. Based on this information, a job applicant may take a certification exam to improve their chances of being hired. Or they may decide to challenge the process, to correct their data or to challenge the process or its result.

Labels should enable comparisons between related products or related processes. For example, they should enable comparisons between hiring practises of different companies, implying a standard. And finally, they should be composable and computable, derived automatically as data and decisions are produced by these complex multi-step systems.

My students, colleagues, and I have been building technical methods for deriving such labels automatically or semi-automatically. But how can we ensure that these labels actually work? How do we explain the meaning of the equivalent of calorie subjectives to algorithmic decisions?

And how do we even know what their concerns are? How do we know what information to surface on these labels? How do we design such labels or other types of disclosure oversight, or, more generally, accountability mechanisms collaboratively and collectively? And for this, I believe strongly that we need to get on the same page. That is, we need public education.

So let me stop here for a discussion. And after this, we will dive right into a description of the public education force that is called We are AI that I hope responds to some of the challenges that I outlined.

ANTHONY MCCOSKER: Thanks, Julia. I'm just going to help direct questions. There's a couple of questions that have been posted in the chat. So I'm going to just point to the first one. So Tanya asks about the reach of the regulation that you are talking about, particularly around the employment

algorithms, whether those questions of jurisdiction are complicated-- I'm sure they are-- and how they've been settled so far.

JULIA STOYANOVICH: Yes, so this is a great question. So let me just say the law hasn't been passed yet. We expect that it will move now that we-- I'm sure that people haven't followed the mayoral elections here, but we just had the primary. And so now we have a sense of who the mayor may be, and so now the city can unfreeze and start actually going about its business. So we expect that there will be some movement on this bill over the summer.

So once it's passed, and also as we're moving towards passing the bill, there are negotiations, part of which is really about jurisdiction, right? What does the city have to say here? Where do we have to make sure that we're compatible with federal regulations?

As far as I know, there isn't anything comparable, either at the state or at the federal level specifically here. We don't have any regulations that pertains specifically to the use of algorithms in hiring. But we do, of course, have various other types of protection for hiring and employment. This is within the purview of the Equal Employment Opportunity Commission that I mentioned earlier.

So all of this is going to need to be fleshed out. But there is, of course, precedent. We have various other laws, including privacy laws, where state and local laws supplement existing federal regulatory and legal mechanisms. So I'm not a legal scholar, and I cannot comment on this further, but I do know that this is really very closely part of the discussion here.

ANTHONY MCCOSKER: That's great. And yes, I think with all of our lawyers in the audience that there's a lot of work going on in and around those questions. I might jump in with a question for myself because I have so many to ask before turning to Jane.

So I had some conversations recently with a person who is developing a startup in Australia that was around automatic hiring systems, but from an employment services perspective. And I just wanted to go back quickly to your point earlier about one of the responses being companies or startups having to make a case for the use of a particular automated decision-making system. And thinking, do you have any good examples, or have you seen any good examples, of that kind of case-making documentation?

Because I could see this person's point that he was looking at using propensity testing, for example, to help people to find the right kinds of jobs that would suit them and they're best placed to do well in interviews, et cetera. So it wasn't using that kind of funnel approach that you were mentioning, so I'm just wondering if you have any of those kind of case-making examples that put the emphasis on what the point or the use of the AI system is.

JULIA STOYANOVICH: Right, so this is a complex question, just like everything else in the space. I mean, one thing that I'll have to say kind of up front is-- I already mentioned this-- is that employers choose to use these tools for whatever reason, but job seekers have no choice. They are subjected to these tools, and whether or not they know it, they are part of the funnel, because for you to apply for a job, you need to know that there is a job opening. And already this is governed by an automated system. The advertisement serving platforms govern this.

So the funnel is there and algorithmic hiring is one of these domains. It's a very clear domain where the promises of efficiency for job seekers, for employers are great, but what are the costs? And how much control do we need to have as individuals? How much control do we need to demand over our own data, over our agency, and how much agency should we be allowed to exercise in these systems?

I think this is kind of the biggest point here. Now companies choose to use these systems for any number of reasons. I think that it's an industry that feeds on itself. I mean, once systems of the sort have been instituted, now it becomes much, much easier to philtre people out quickly. Now every employer is compelled to use a tool of this sort because they can no longer manually sift through all the applications. And so it snowballs.

But again, algorithmic hiring tools are one of these systems where we need to ask the question as to whether we need them. I mean, what is it really in society that we're trying to automate? Is there actually a gain for anybody involved? Or are we just using these tools because they exist, someone invented them?

I'm not sure what some of the positive examples might be for a company. We are now starting a project, as a matter of fact, with Mona Sloane and with another collaborator, Tessa West, who is a social psychologist at NYU, to look into how these tools specifically are used by organisations. Because the only thing that we know from the outside is these tools exist, and you put in some data, and out comes a result that may be arbitrary, or it may be gender biassed, or it may be racist.

But we don't know to what extent hiring managers trust these tools, even. Are they actually playing an important part in the entire cycle of this human-machine interaction? As far as we know, there isn't research on this. If any of you in the audience are aware of such research, we would very much appreciate pointers. But I agree, Anthony, that it's important for us to understand how people actually use these tools to know to what extent to be concerned. And I expect that there is no uniform way, that it depends on the company, depends on the context.

ANTHONY MCCOSKER: Yeah. Thank you. Jane, did you want to jump in and ask your question before we get back to--

JANE FARMER: Yeah, actually, I think you might be about to answer my question, so I did have a different one.

ANTHONY MCCOSKER: OK.

JANE FARMER: My question is, is regulation the only solution? So I guess when you talked about pre-existing technical and emergent bias, I thought all research could be open to that. But we have ethics committees and human research, and technically that's supposed to stop that happening.

So are there other approaches, I guess, is what I'm saying. Should there be more ethical employees like the algorithms management and policy officer? Could that be an ethics kind of job?

JULIA STOYANOVICH: Yeah, you'll never hear me say that one size fits all. And if you ever do, then please remind me that it's time for me to go do something else or retire. Absolutely, regulation is

not the only thing that we can do. It's one of the tools in the toolkit. And the reason that I kind of used regulation here as one of my motivating examples is because I've had, like I said, kind of unexpectedly I've had some firsthand experience with it, and I just wanted to share how that all went.

But you absolutely can motivate public education and public engagement work from any other angle. And really the way that I think about this now is that we need to strengthen our shared accountability, distributed accountability structures. Everybody's responsible. I'll show a picture that shows this. Everybody is responsible for making sure that we keep these systems in check.

We can say it's just the evil companies, or it's just the evil employers, or it's just the government's job to oversee. Because many of the things, many of these unintended consequences of the use of these systems, only the person who experiences them knows that they need to surface. And the example that I like to use there is somebody with a disability. Disabilities are so varied. By checking a box saying I am or I am not disabled, you are not giving enough or a lot of information at all to that system.

So you cannot audit for whether the tool discriminates against individuals with disabilities because it's such a heterogeneous group. And there are all these intersectional effects, the disability compounded with other demographic group membership and with other disadvantages really brings. So I think that really it's up to each and every one of us as data subjects or individuals being impacted to keep these systems in check. And this is why we need to be educated and to understand how to speak up.

ANTHONY MCCOSKER: Should we keep going with your side of the presentation, Julia?

JULIA STOYANOVICH: Yes, let me share again, and please confirm that you don't see the notes. I don't see you, I only hear you all. So--

ANTHONY MCCOSKER: It looks good

JULIA STOYANOVICH: All good?

ANTHONY MCCOSKER: Yeah.

JULIA STOYANOVICH: OK. So now I'll speak about a course that we developed called We are AI. This is a collaboration between the Centre for Responsible AI at NYU, P2PU, which stands for Peer to Peer University-- and that is a public education nonprofit-- and the Queens Public Library in New York City. So Queens is one of the boroughs, and the Queens Public Library serves that borough.

Course materials were developed by me together with Eric Corbett whom I showed earlier on-- he is our human computer interaction and public engagement and public participation expert-- and also with input and participation and support from Mona Sloane, Falaah Arif Khan, Megan McDermott-- that's all from the right side-- and then also Becky McGraff, Chris Peterson from P2PU, and from Queens Public Library, Jeffrey Lambert, Sadie Coughlin-Prego, and Kevin Bora. So as you can see, this was a large team.

Materials for this course are all publicly available on GitHub. Everything we do is on GitHub. They include a series of videos, a comic book series. And both of these are common fives, so there are five volumes to the comic book series, and there are five videos, one for each of the five modules of the course.

Anyone can use the material that we provide to facilitate what's called a learning circle. So what's a learning circle? It's a group of people who come together to study some materials together. You can think of it as kind of a generalisation of a book club. And in particular for our course, to follow it, no math and no programming skills or existing understanding of AA are required.

As I just mentioned, the We are AI course is designed to be run as a learning circle. And this is a methodology that was pioneered by P2PU. A learning circle is a facilitated study group for people who want to meet regularly and learn about the topic with others.

There are no teachers or students in the learning circle. It's a group where everyone learns the material together. The facilitator decides on the meeting schedule. They keep the group on task, so they are kind of a manager, and they support individual learners' participation and goals. And all of the materials, group prompts, and activities need to run as a group with minimal preparation. And they are all incorporated into the course. There's also no homework here, so all work takes place during the meeting.

I will now highlight each of the five modules in this course. And like I said, each module has an accompanying comic book. I'm showing the cover here on the left. And on the right, I'm showing the structure of the module.

So module AI-- sorry, module one-- is about defining AI. And as you may know, no agreed upon definition of AI exists today, so I start by giving my own definition. And that is an AI is a system in which algorithms use data and make decisions on our behalf or help us humans make decisions.

And while AI may be difficult to define, we may hope that we'd know it when we see it. So this is actually one of the competencies that we teach people is to recognise AI. And here I'm giving some examples. I showed a Roomba, a robot that helps us clean. I showed chess-playing AI and a smart light AI, and that is an AI that turns on the light or turns it off on our behalf.

And here are a couple more controversial examples of AI. And that is a self-driving car and an AI that helps assess people. For example, interview candidates for a job, no surprise there, or assess credit worthiness of applicants.

And we talk about the following, that to start understanding AI, we really need to discuss each of its components and that these algorithms data and decisions, and then we dive into that. We consider the task that many of us attempted with mixed results during the COVID-19 lockdown, and that is baking sourdough bread. And this is a task that I use to explain algorithms, data, and decisions. I give a recipe.

And I say a recipe with these steps, it's an algorithm. It lists the steps that we take to transform the ingredients into a loaf of bread. The algorithm may be fully proscribed. We call them rule-based.

So such algorithms, rule-based algorithms, list exactly what ingredients to get, how much of each to take, how to mix them, and the temperature to bake. If we know the rules well enough to write them down, then we can always bake a loaf of bread. But sometimes we don't know the rules, and we need to learn them from data.

And we also may want to see some variety of the outcomes, and this is where these learning algorithms come in. And here I talk about learning algorithms that learn the recipe to bake sourdough from our experience of what good sourdough tastes like. And then here we also go into different types of data, such as inputs, parameters, outputs, and human judgement.

And the comic and the video for this first module, defining AI, ends on a quote from a wonderful recent book called The Age of Algorithms by Serge Abiteboul and Gilles Dowek. The goal of this punch line is to reinforce human agency. And the quote here is that "creations of the human spirit, algorithms," and AI-- and I'm paraphrasing here by adding and AI-- "are what we make them. And they will be what we want them to be. It's up to us to choose the world we want to live in."

And this theme that technology is not just technology, it's not just subjective, but there is always human judgement, there's always everyday intelligence, and we always must exercise agency in our interactions with the systems, is what runs through the course. The second module of the course is the most technical, and it speaks about learning from data. So here I'm showing an outline on the right.

And the way that I set this up is suppose that you are asked to design a smart light system, an AI that decides when to turn on the lights in your house and when to turn them off. How should your AI decide when to take these actions? And we can start our design process by postulating a simple rule based on our own everyday intelligence. We call it rule one, and there's going to be a rule two later on.

And that rule says turn on the light if it's dark outside, otherwise keep the lights off. And here we remind people that an algorithm is a sequence of steps, but this algorithm is just one step. It's very simple. It's just this one rule. The input to the rule is whether it's dark. The output is whether to turn on the light.

And then crucially, we talk about prediction accuracy. So the rule is simple, and simplicity is good, but is the rule any good at predicting what it means to predict? Does it appropriately capture how you would manually operate the lights in your house?

To check whether it's any good, we would run an experiment. We collect observations about whether it's dark outside and whether the lights are on in the house. For each observation, we will check whether the prediction made by the rule in fact matches what we observed.

And then we go through this process. I illustrate how you actually collect observations. And in summary, we can see the performance of this rule that was correct three out of four times, and so it's predictive accuracy is 75%.

And then we talk about some terminology. We discuss inputs and features. We discuss outcomes and labels. We discuss classifier rules.

So the rule that we showed here is a classifier, and it makes one of two choices, turn on the light or turn off the light, and so it's a binary classifier. And then we talk about the cost of a mistake. So this was an observation where the classified said to turn on the light, but it would have been kept off. What is the cost of this mistake? It's probably that your smart light woke you up in the middle of the night.

And so what do we need to do now? We need to decide to collect additional features because based on the information that we have, we actually cannot tell when to turn on the light and when to keep it off, so the rule was too simple. And then we're thinking about what features to collect. And this is crucially important.

If you go back to algorithmic hiring now, all these companies are pretending to predict someone's performance on the job based on their shoe size and on whether their first name is Jared. So these features are irrelevant, and people absolutely need to intervene here to actually think critically about whether we have any reason to believe that the feature is useful. And I'm illustrating this here already with my smart light classifier.

What features should we select? The outside temperature, probably not. The price of tea in China, definitely not. Now what about whether it's bedtime? And this makes sense based on our everyday intelligence.

And here we design another classifier, and it worked well for us. But now we're going to go and try and use it at another person's location, at an office, and we will see that it makes mistakes again. And this is because we were generalising out of sample here.

So I'm not going to dwell on this here, but the point here was to show that in something like 12 minutes, in the second module, it's actually possible to convey technical material without formulas and to explain the limitations of observational data, to explain uncertainty, to explain that mistakes are associated with costs. And whose mistakes we prioritise, which mistakes we want to lower, is going to then be beneficial to some stakeholder groups over others.

And then we talk about learning from data. How do we learn these rules? That by observing the data, we can actually figure out a lot about individuals, about when the holidays are, for example, in their countries, and this is why it's dark in the offices in which they work. But ultimately, even if we learned our rules from data, we need to validate them with the help of the scientific method.

And we end this module on the importance of formulating falsifiable hypotheses so as to test whether your classifiers, or more generally your algorithmic systems, work. The crucial question, does the classifier work, is formalised as, are the classifier's predictions more accurate than a random guess would be? And this is the lowest bar for accuracy. You don't need the fancy AI to be flipping coins.

When we have ground rules information, and that is whether the light indeed should be on or off, we can check. And if the hypothesis is falsified, meaning the classifier doesn't work, it's no better than a random guess. We should be prepared to redesign it or not use it.

And then we end here on the fact that the classifier still will be making mistakes, and this is because prediction is difficult and often impossible because of the uncertainty in the world, because the rules are sometimes broken. And from here, we go into module three that talks about ethics head on. And the title of this module is who lives, who dies, who decides?

The illustration is of a trolley problem serving as a can opener for ethics. And really, this comic and the video that accompanies it, they are critiques on the trolley problem. They underscore the problem with the trolley problem, or the problem with algorithmic morality.

I'm actually going to probably skip this because we're running out of time, but I encourage all of you to look at the comics. Here we cite Jeremy Bentham when we talk about algorithmic morality, how you can't measure happiness, and how you can't outsource the work of being human to a machine, and that we need to make technology rooted in people by considering ethics head on. And there are two more modules here.

The fourth speaks about bias. Here we return to algorithmic hiring systems. We talk about data as a mere reflection of the world and the limitations of what you can glean from observational data. And then the final comic, we really enforce and final module, really enforce agency, that it really is up to us to decide whether we trust the data, whether we trust the algorithms, whether we are comfortable automating the decisions we're automating, and that we must check that the tools actually work before we use them.

The final point here is that there is a shared accountability, that we all are accountable for the decisions that these systems make. And to wrap up, I think that our goal here-- and this is what I try to underscore-- is that we should be building systems, and we should be helping members of the public understand and participate in the oversight of these systems. Our goal should be to have a nuanced conversation about the role of technology in society that goes beyond techno optimism-- and that is a belief that technology can fix deep-seated societal problems like structural discrimination in hiring and techno bashing-- this is shown on the right-- that believes that any attempt to embed ethics and legal compliance into tech is just fair washing and should be dismissed.

And I think that for me as a technologist, as an educator, the main responsibility is to help build systems that expose the knobs of responsibility to people. And I'm going to end here. This is another comic book series that we are working on that is for a more technical audience, so I encourage you also to take a look at this one. And that's it. Thank you very much.

ANTHONY MCCOSKER: Thank you so much, Julia, amazing material. And I'm just sad that we don't have more time to discuss. And I'm conscious that many of the people along today might have had things booked in at 10:00, but I just want to say one thing quickly, just to underscore your point about looking at the material on GitHub, because I think it's material that we can continue the conversation around. And given that you're part of the Centre for Automated Decision Making in Society, we hope to continue these conversations as we go in various venues and forums and workshops and meetings.

So it was really amazing to get an overview of your work, which is really extensive. It's really at the edge of, I think, what the whole Centre was put together to do, as well as our work in Social

Innovation Research Institute at Swinburne. I wanted to also just give you my love for the visualisation side of things, because I think there's so much that can be done in bringing together data science, computer science, and social science and humanities when we use art and illustration and conceptualise issues, processes, and technical material through visual medium. So thank you for that.

And I just want to also let everyone know that our final webinar in this series will be on the 21st of July with Amir Aryani, and that will be about community data collaboration projects that we've been involved in. So we hope to see you there as well.

So thank you very much, and enjoy the evening, Julia. And--

JULIA STOYANOVICH: Thank you very much. Thank you for having me. It was such a pleasure. This is the first time I gave this talk, and that's why I guess I'm so enthusiastic about it, but yeah. Thank you for having me.

ANTHONY MCCOSKER: Yes, and--

JANE FARMER: Thanks, Julia.

ANTHONY MCCOSKER: Don't forget, everyone, like and subscribe to the YouTube channel so that you can watch the recording. And as I said, we'll continue the conversation. Thank you.

JULIA STOYANOVICH: Thanks. Bye-bye.

JANE FARMER: Thank you.

[END OF TRANSCRIPT]