Welcome to Disrupting Disruptions, the Podcast of The Feminist and Accessible Publishing, Communications, and Technologies Practices Speaker and Workshop Series. Today's episode features Dr. Sarah Myers West of the AI Now Institute, speaking about AI and Intersectionality.

This series seeks to bring together scholars, creators, and people in industry working at the intersections of digital humanities, computer science, feminist studies, disability studies, communications studies, LGBTQ studies, history, and critical race theory. The series will bring forward critical approaches to publishing practices, communication strategies, and techniques for making research dissemination more accessible.

Part of the motivation of this series is that while humanities and social science scholars will critique "traditional" academic publishing and communications strategies as being sexist, classist, racially biased, and inaccessible, the kinds of solutions proffered (such as open access and allegedly "innovative new technologies," etc) often romanticize and fetishize technological alternatives and do not look at how inequity can be perpetuated or only shifted (especially at the level of algorithms!). The series is organized by Dr. Alex Ketchum of the Institute for Gender, Sexuality, and Feminist Studies at McGill University and all events have taken place in Montreal, Canada on unceded Kanien'Keha:Ka territory.

This podcast makes our events accessible to a larger audience. Full transcripts and video links are available at our website, disruptingdisruptions.com

Today's episode with Dr. Sarah Myers West was the 9th event of the series and took place on September 19, 2019.

Dr. Sarah Myers West is a postdoctoral researcher at the AI Now Institute of New York University and an affiliate researcher at the Berkman-Klein Center for Internet and Society. Her research centers on the critical study of technology and culture, with an emphasis on historical and ethnographic methods. She is currently working on a project that addresses the politics of diversity and inclusion in technological communities by exploring the nexus of artificial intelligence, gender, and intersectionality. She received her doctoral degree from the Annenberg School for Communication and Journalism at the University of Southern California in 2018, where her dissertation examined the cultural history and politics of encryption technologies from the 1960s to the present day. Her work is published in academic journals such as New Media & Society, Policy & Internet, Business and Society and the Internet Policy Review.

———

Sarah Myers West: Thank you so much for having me today I honestly feel really honored to be a part of this speaker series to be in the community as well as advocates with so much of my work and understanding of this, this area and thanks very much to Alexandra for putting together this series to all of you for coming tonight, really fantastic.

So my name is Dr. Sarah Myers West

I am at the AI Now Institute, and I titled my talk today discriminating systems examining the social implications of artificial intelligence. And I want to take this, I want to start by taking a step back to talk through what we actually mean by artificial intelligence, a word that kind of gets thrown around a lot. It's very fuzzy and to bring us all into the space of understanding. So often when we think about AI, these are the images that pop into our heads. They're machines that are designed to replicate human intelligence. And ultimately, we do something that's kind of along those lines in devices like Siri, and Alexa, but no these systems are still pretty unsophisticated. I don't know if any of you have talked to Siri recently. But they're really nothing close to what the movies tell us that AI is. And in some ways that might be reassuring. AI as being used to the world around us doesn't really look very much like human intelligence, and in fact it’s folks like Meredith Broussard have pointed out, often these things are fairly unintelligent most AI systems as used today rely on applying statistical methods to very large amounts of data, looking for patterns in order to make predictions about future events. And this is how AI is currently being deployed across a wide range of domains and often in ways that are much less visible. So, when we buy something at a store for example, artificial intelligence might be at work in determining the price of the goods on the shelves, or assessing what needs to be restocked is used to evaluate valuable

products for buying to offer as coupons to try and get us back. Maybe at work for example in shaping the traffic patterns as we drive home. And then telling us the optimal routes for how we get there. And those are all pretty mundane examples, but we also see AI being used in ways that are much more deeply consequential in healthcare settings AI might be used to make recommendations for cancer treatments. It might be used to determine who receives Medicaid benefits. It can determine where our kids go to school, it can determine whether or not someone will go to jail for their sentence, let me dive a little deeper into what this looks like.

So it was recently revealed that Amazon has developed an internal hiring tool that would help the company to rank and assign scores to resumes that candidates submitted for open positions. The idea was, they could train the system to help it surface the best candidates they receive an awful lot of resumes, every year. And then they were looking for who would be the best fit for Amazon based on data from the past hiring decisions, kind of similar to how Amazon makes recommendations, where you know what book you might want to buy next based on your browsing habits. The goal was that through the system they surface the best talent and also try and reduce the likelihood of bias found was that this system learns to downgrade candidates that attended certain universities. And in fact, to

downgrade others for even mentioning the word woman in their resume. Now, the engineers developing the system tried to apply techniques to reduce the bias but they found that it was basically baked too deeply into the system for it to work. They couldn't develop a functional system without discriminating and ultimately they ended up scrapping it entirely. What does this example teach us? For one, it illustrates a principle of garbage in, garbage out retreating data that Amazon views was based on past highs. And it probably comes as no surprise that the company's engineering workforce is made up predominantly of men. This is why qualities associated with women down ranked in the system. It was identifying patterns and replicating those patterns. But it also shows how difficult it is to build AI systems it’s a really difficult, complex problem. And with some of the world's leading machine learning researchers and after applying, you know, cutting edge techniques and data science. They can’t produce a system that wasn't discriminatory so what does that sort of leave the rest of us to do.

So at the end AI Now Institute we’re studying the consequences and social implications of AI across a wide range of domains we have an interdisciplinary team of researchers working in hiring on education on the environment on healthcare. Security criminal justice is a hub for the emerging field of researchers

focus on these issues. We're the world's first Research Institute that's devoted to studying the social implications of AI, and we've remained the first and only AI Institute to be founded and run by two women. Today it's maybe worth talking briefly about how we are approaching this research. I really strongly believe that in order to understand this field, it's really critical to draw on all of the methodological tools available to us. And in particular, we believe that the social sciences do have really critical roles to play in helping us make sense of our technological culture. And so, although AI is often positioned as a STEM field, and the field, history, actually is a very rich interdisciplinary crossroads of psychology of behavioral science, anthropology and philosophy. And we also see in processes through practice the erasure of the contributions of women trans people and people of color. And so in the work that I'm presenting today I think it's surfaces the need to render more visible, This is work in the history area that this is actually a tool to the field. And I think that there is a clear need for scholarship in this space. And just what I'll be discussing. And this is a pilot study that forms the first year of my postdoctoral research at AI now. And it points to the background of a larger research program that I'm meeting in the dynamics of gender, race in power in AI. And I'll be honest, when I first approached doing this landscaping study. I expected that I wouldn't find that much I thought that I was going to find gaps. And what I

found was really a very rich, historical being of scholarship that for decades identified these, these issues in the field. And so, we read over 150 peer reviewed studies. But I also really want to foreground that knowing that there was a history of erasure. We wanted to value informal knowledge creation that the ability to publish all of the work is an area that can be policed. And so we wanted to also look at informal knowledge, and so I found blog posts like medium post by people working in tech companies can be a really rich place to go to understand what's happening, and especially in the present day. I do reports, I looked at press articles and found that to be a really rich space.

And. And what we found is really that, AI in the midst of a diversity crisis and you might be familiar with the diversity challenges in the tech industry overall. In 2013, the share of women in computing dropped to 26% and that's below the levels of representation of women in the 1960s. So in 2013 it’s lower than in 1960 levels are almost half of women that go into technology, currently eventually leave the field. That's more than double the number of departments. And we found ample evidence, spanning decades that things are much worse in AI than in computer science overall. We found in a handful of the leading AI conferences 18% of authors are women. 80% of AI professors are men. Women only comprise of 15%

of AI research staff at Facebook currently 10% At Google, we don't have any public data currently available on the trans community fields.

Things look even worse when you take race into consideration. And so at Google, the highest rate of attrition among workers is among black employees at the percentage rates are pretty small across racial categories, but there's way more to the story than what statistics can show us. And you see this crisis as it unfolds the process of all of the biggest players that are involved in developing AI today. From a class action suit for the by Microsoft workers alleging a system, a systematic failure to take allegations harassment and discrimination, to a federal investigation into gender discrimination at Uber to Apple’s concerns about its lack of workplace diversity issue, while also at the same time saying that proposals for adopting diverse hiring practices would be, for instance, and about pay practices about six to seven standard deviations between pay for women across almost every job category. And we'll get into it and talk about the phenomenon. Black employees that Facebook have recounted being aggressively treated like by campus security and dissuaded from taking part in black group activities and at Tesla those alleging gender discrimination and a hostile working environment where one employee

recounted that there were more people named Matt in her group than there were women.

So, clearly, we know there's a problem and the evidence is not really the issue. And in fact, in doing this work, I found mountains of studies that were focused just on the question of the pipeline. And this is a term that's used often in industry to refer to the absence of diverse candidates in the hiring pool. It's often used by friends to justify why they can't hire. So here are a few examples of what that literature looks like. And you see titled like ‘why are there so few women computer scientists’, ‘where have all the girls gone’, ‘what draws women to and keeps women in computing’, ‘why do some gender gaps remain while others do not’ and my personal favorite, ‘will computer engineer Barbie impact women’s career choices’. So, these studies rely predominantly on survey based research that's conducted in educational settings, almost always in universities, University undergrad classes, universities, on coasts, in the US that are predominantly elite universities, and they try to understand the factors that lead to gender based discrimination in computer science, more precisely by trying to interrogate what drives people in to it, and implicitly stay.

And I also want to note that these studies, largely treat gender as a binary phenomenon which will erase the experiences of members of the trans community. And they also don't acknowledge the ways in which oppression intersect, that people are impacted differently. When you also take into account differences of race possibility among other characteristics. And what this means is that all of the women and tech initiatives that come out of this implicitly benefit white women over all others. So there's various work being done in the way that these studies are positioning. Although, certainly, like, they're an important start. So they center the role of culture and suggest that a student's self assessment of whether they're a good fit with the field is likely to influence whether they leave computing, And it's intertwined with stereotypes of computer scientists and singularly focused a social candidate as white, single and male. And they assert that women tend to persist in computer science when they reject and find alternatives to the dominant culture.

And so, although they make a contribution that helps us understand the factors that influence participation in technical fields. There are a lot of limitations that go unacknowledged in the way that these studies are taken for one we rely on samples that are often series that are conducted in classrooms settings, sometimes for the professor who is leading the class, which could have an influence on responses to

surveys and their self reported data, and they often have pretty small sample sizes like 30 or so people from one university in a particular place and time. And they also place the onus for change on those who are discriminated against and, and then raises the question of what's the work that these studies are doing. They're a fairly narrow frame to which to view through which to view potential variants. And they largely address the solutions toward educators, and although that's certainly a critical gateway, it’s not an excuse not to address discrimination in recruiting environments and in the workplace. To address actually existing racism, ableism, misogyny. And so, in the report we argue that diversity initiatives really need to be accompanied by efforts to address workplace cultures, and the logics of how technical systems are designed and cultures of exclusion that have been very frequently documented, but that remain unaddressed I remember seeing one study that looked at the philanthropy initiatives. I think it was 27 major tech companies. And they found that 0.4% of all philanthropic giving in 2017 went specifically towards initiatives geared to women of color. That was really $350,000 across 27 tech companies, and that’s like a drop in the bucket when you look at company's profits. So despite the volume of these studies, despite the relative consistency of their findings we really don't see that substantive change.

And in fact, you can see similar trends historically. So, you may already know that computer programming was originally seen as women's work. And it was when programming began to be seen as a professionalized and expert domain it was gendered male. As Mar Hicks writes throughout history it has often not been the content of the work, but the identity of the worker performing it that has determined it's status.

Let's take a step back for a moment. What I just traced here is that discrimination in the field of AI is expensive, there's ample evidence to support it and that has a very long history. But I want to shift gears to talk a little bit about the consequences as we see them emerging in AI technologies themselves. So I started talking earlier about digital voice assistants like Siri and Alexa, in a report that was released recently, it talked about how these assistants are perpetuating gender biases because of the way that we relate to these devices. These devices tend to have feminine voices on default, sort of position our relationship to them, in which women are situated in a position of servitude, and that the micro interactions that we have on a daily basis accumulate over time, in ways that will perpetuate gender biases. And if you look across the entire product category nearly all such products are gendered feminine. And this is a particular cultural choice these companies

have like very lengthy biographies of like, who Siri is or like Cortana if you look at the BIOS for how Cortana is vision she's like, the child of two professors, one of whom is like a historian and and the other one economist, and she grew up in like New York but then she's traveled all these other places. These are really rich, rich, like imaginaries that are developed around these systems, and they're very particular. And if you look across different countries. These defaults don't necessarily hold. So there are countries in which the feminine default is actually masculine and that's in relation to gender. We might consider. Well, why is it that this is the default that we see in the vast majority of the market.

Well for one this is a perpetuation of gender dynamics and have existed for a really long time. The earliest test of voice recognition software, literally couldn't hear women because they were developed in labs where the voice that they were testing on were largely the voices of men. Another computer program that you might have seen this this study by propublica, if you haven't, I really encourage you to take a look at the story, and it illustrates the racialized dynamic you see in the systems. And so, compass is a computer program that was designed to predict the likelihood of recidivism. To determine if someone who is accused of a crime or who committed a crime is likely to commit that same crime again. And it was used by

judges to determine the amount to bail at because you want to set a higher bail for someone who is going to go in again. And propublica found that not only did the compass exhibit racial biases, but these racial biases were very frequently wrong. And it was both an ineffective system. And it was also a discriminatory system.

These concerns also have important implications to our safety on the roads. So Uber's testing out of using self driving vehicles at several locations in the US and in 2018 one of these cars hit and killed a pedestrian. This is, as you can see from this image, and the pedestrian was wheeling a bike across the road, and the computer vision software didn't really know how to make sense of this image, you know, it uses, it looks at the lines on the road. But it had trouble recognizing this one because it was an image that it wasn't hadn't encountered before. And the scroller here and this describes how this case study is really concerning for members of the disability community. Because what this image looks a lot like autonomous vehicles have trouble recognizing bodies that move differently in the world. And that's a deep safety concern.

So I can give you many many more examples of discriminatory systems. So dispensers that can't see skin with darker skin pigmentation to cancer screeners that have higher failure rates to detect cancer in black patients because they're

trained on data sets of predominantly white patients, and in some of my ongoing work, I'm looking at the development costs. I'm positioning as a class of computer vision systems that try and make inferences about your innate qualities on the basis of your physical appearance. So things like race and gender recognition systems that are developed with very specific ideas of what race and gender are. To affect recognition systems which, some of which try and assess your mood or your personality traits on the basis of your facial expression. And these systems are already in wide commercial use they're already being used to make decisions of great importance every day.

But what I've traced out for you so far is a feedback that discrimination and inequity in tech has really significant material consequences, particularly for underrepresented groups who are already excluded from resources and opportunities.This reason alone is should be enough (inaudible). But in the case of AI these patterns of discrimination exclusion reverberate beyond the workplace into the wider world.

And that's why I want to bring another concern, which is that the remedies that are being proposed to discriminatory systems risk amplifying harmful practices instead of solving them. So the solution to bias AI is proposed as often an issue of

diversifying the dataset, the endgame with that will lead to proposals that result in increased surveillance of the very communities it most harms. So for example, Uber drivers have to often show their like take a photo of their face to authenticate themselves. It’s a security measure to make sure the person driving the car logged into the account are who they say they are. This system has a really difficult time recognizing the faces of trans people. Lots of people were getting locked out and they'd have to drive, hours to go into person to log into their accounts. And I said, a researcher is trying to solve this problem, went and scraped the data from YouTube videos of people that were undergoing transitioning, and that's a deeply personal experience to go through and that was done without consent of the people involved For their data to be used for that kind of use. So, we see an amplification of, you know, practices of surveillance in order to solve these other kinds of discrimination problems that often, then go on and are implemented in systems that surveillance themselves. So, rather than to play quick fixes we should be asking questions like what kinds of assumptions about workability and potential become coded into these systems and who is at the table for when they were built.

So this is the focus of the next stage of my research. This is a very neat presentation of a really messy phenomenon. It's not just like a one to one cycle.

There's a lot that goes on in this middle space, and that’s a space for interrogation. Interrogation that I think social science and humanistic approaches to research can have a lot to contribute.

So I'm looking at how AI is reshaping how we understand labor and reading this through the lens of race, gender and ability. A project still in it's pretty early phases, but I want to highlight one study that was published last year looking specifically at the use of predictive hiring tools across all phases of employment. So, this ranges from, how job ads are placed, who it gets targeted to. Are you being shown higher paid jobs or are you being shown lower paid jobs, and to the scanning and raising and evaluation of resumes like the Amazon example that I started with. The use of tools in video interviewing that will listen to things like your ums and your pauses and your vocal fry in order to make assessments of what kind of job candidate you are. That’s in wide use by the way there are some banks that are no longer conducting on campus interviews, and only using video interviewing in order to use these tools. Some make recommendations about what salary are going to be offered, and they found significant biases. One of my favourite anecdotes is they talked about a resume screening company and found that it’s model has identified that having the name Jared and having played High

School lacrosse is going to be a very good predictor of success in the job, which purely had nothing to do with any like no positive attributes of the job. But what it does is it perpetuates certain ideas about culture fit, you know, who is going to be a good fit for the company. And, and that is going to amplify existing discriminatory effects, and sometimes in ways that are a lot harder to track and make sense of. Most significantly, these biases surface even whens schools explicitly ignore race, gender, age, and other protected attributes because this is an area where we have a lot of anti discrimination law. But even when you exclude them you see them emerging through proxy variables, things like your zip code, things like your name can replicate these patterns of inequity. Even when the kinds of things that are protected under law are taken out of consideration.

So throughout this talk, I've highlighted the ways in which human's do remain infinitely more sophisticated than what these systems provide, but we are still deploying these systems at large. Even though, as though they're going to be fair or more objective than what humans do. So, one thing that I'd like to leave everyone with is the need to be wary of automation bias, which is something we're all prone to this idea that we give outsized weight to the outcomes of automated systems and despite the expertise that we all hold. That's not to say that we aren't flawed, but it

is much more difficult to understand flaws at scale in the way that these systems are built and also the way that trade secrecy is used to make it very difficult to scrutinize it from outside. We also shouldn't give weight to claims that this our inevitable. So, regulators are looking at a variety of forms of algorithmic accountability in the US, the algorithmic Accountability Act for Congress that we require companies to assess whether there are bias or discriminatory outcomes of algorithmic systems. And we see discussions about banning the use of facial recognition here in Montreal among members of the city council. We've seen bans enacted in cities throughout the United States. And among tech workers themselves in the face of inaction by their employers. Tech workers are walking out and trying to construct the kinds of workplaces that they want to take part in ensuring that they're inclusive of everyone regardless of race, gender ability, sexual identity, class or importantly contract status, and many of these people in Google over half of its employees are employed on contracts on their very precarious positions, refusing to take part in projects that they see as immoral. Tomorrow in front of the US we see the first walkout by, I think now there's eight major tech companies, that are going to be walking out for the environmental impact of what their companies do. The first time we've seen workers from across companies come together.

So, we see the tech industry as a harbinger for things to come in the future of work. Both as you know, a space for possibilities of hope but also a place where we really have a lot of hard work to think critically about what it means to position folks that are working at tech companies as the locus of change, there's reason to think critically about that. And so as students as researchers and educators think everyone in this room is uniquely positioned to play a critical role in shaping what that future looks like. And lastly, and I would be remiss if I did not foreground some of the studies that deeply shaped this work. These are a few that would be a really fantastic place to start. If you go on the AI NOW Institute's website we've also put together a playlist that came out of the initial research that we did that highlights some of the historical work, and you'll be able to find that on the website (https://medium.com/@AINowInstitute/gender-race-and-power-in-ai-a-playlist-2d3a44e43d3b) . Thank you very much.

Alex Ketchum: And also that playlist that Sarah just mentioned is also linked to the series website (https://www.feministandaccessiblepublishingandtechnology.com/p/reading-lists-we-love.html) as well so I think we have some time like 15 minutes or so for questions.

Question from the audience: So thanks for that talk. I'm actually interested in hearing a little bit more about AI Now as an institute, like as we know there's groups like the ACLU, or ESF taking equally adversarial relationships to fight for civil liberties. And I just be interested in hearing a little bit about how AI now thinks of itself as an organization, and how it thinks of itself as being able to kind of catalyze change? Aside from research, obviously, yeah, which is something important, whether they're kind of other modalities at play.

Sarah Myers West: That's a really good question I'll be honest, that's a question that we're kind of working through day by day. I've been with AI Now for a little over a year now as an institution we've only been existing for like a year and a half coming up on two years. And, and we're definitely research first. We are a research institute, but it's research that's that's devoted towards creating social impact. And obviously, folks like Meredith Whittaker who is the co founder of AI Now, who is integral in a lot of tech work or activities that we saw at Google and elsewhere. So, I think it's very much part of the ethos of the Institute in terms of what it looks like in practice. I think it's a work in progress and we're thinking through the ways that we can inform change.

Questions from audience (paraphrase): Can you speak on the facial recognition bans?

Sarah Myers West: Okay. I can't speak specifically to Montreal's like the, you know, I'm sure that there's a lot of highly contextualized aspects to the way that's being discussed here. There's a number of different reasons why you might want to ban facial recognition one is the ways that facial recognition systems are deployed are often in areas that have significant social impact. And they also have with, I should say, without much accountability for theses systems. They have pretty low rates of accuracy, particularly for darker skinned women, and so if you're using them for example, in a security context, it means that they're going to have disparate impact on certain communities over others. There's also some real privacy considerations. And so, it incentivizes the widespread climate of surveillance and adding to that decision making. There's a number of other really gnarly issues at the moment we don't have any real form of accountability. The first step has been to institute (inaudible).

Question from the audience (paraphrase): You said it's problematic for smaller communities, things like not having a lot of visual data for trans people. Is there any better way to do that? Something that's a little less intrusive?

Sarah Myers West: So I'll be clear obviously this is a really gnarly problem, it's really complex it's really hard. Also, I think, as the, as I highlighted as well. diversifying the data set doesn't necessarily fix the problem. And a lot of the issues with discriminatory systems have to do with the system design itself, the model itself, the kinds of inferences it makes. And so I'll give you an example, gender recognition systems. One of the primary use cases for gender recognition systems, is that it cuts down on compute power, like, especially if you're trying to recognize a face if you can. First, they identify the gender of the face. You can cut out half of the data they got to run through. But implicit in that model is that gender is a binary phenomenon. And that fundamentally erases the experience of folks that do not identify with the gender binary, then you look at how it's deployed in practice. For example in security context and it marks certain bodies as being higher risk. There's all kinds of points at which discrimination occurs that is beyond the data So that's reasons that that's kind of a flawed approach because it's a very narrow--it's really a broader social problem. And two is the issue of consent. And this is a really difficult issues because of a variety of reasons you already pointed on like the economic challenges, it's very costly to produce datasets and so there's an incentive to try and, you know, use train models over and over. And there's a lot of incentives to secondary uses, we haven't had a broad social conversation about what consent

to be a part of a dataset looks like. And, you know, you might have consented to the authentication example, but that data set could then be used to build, you know, for military applications that you don't consent to. And so I don't think that there is a quick fix answer. I think that it's, it's a much deeper challenge. One place that I think you might look to for a generative model, Illinois has an Biometric Identification Policy Act that not only puts the onus on those using biometrics to take into account the privacy considerations of those people that the biometrics are being used, but also provides near term and long term ability to revoke consent for your data to be a part of a data set. That also leads to different kinds of company practices and processes and different kinds of design to technical systems to make that possible. There are other downstream social impacts that having consent in mind (inaudible).

Question from the audience: In the first study that you did the review of diversity in the field. You mentioned a lot about pipelines, and I'm sort of wondering if there was anything you came across about affinity groups like these are becoming more and more popular within AI right now. Lots of these affinity groups are like popping up at AI conferences and stuff like this. I know WIML has been around for

like a long time, but I don't know if there's been any research on like whether or not there's some sort of effect of those if so what type of effect.

Sarah Myers West: I have not seen that research, either. And I should be clear that the vast majority of the pipeline literature is focused on computer science overall, and not specific to AI. I had to really dig by looking at particular programs to get any. So, you know no I don't I don't think that we have very much there. I know that we're seeing the growth the, you know, the screws, black in AI being another NuerIPS, the neural information processing symposium one of the flagship conferences in AI. We're seeing growth of the affinity groups there so I think it's a good space to watch.

Questions from the audience: You mentioned that with algorithms with race and gender they were still discriminating because of the like proxy variables, can you explain more about that?

Sarah Myers West: Yeah, sure. Um, so, proxy variables are essentially kind of a stand in. So you might be excluding demographic categories by race, say you have a number of different loop holes for a piece of data and you might be excluding race from the model. But you can still see racially discriminatory in-hacks that result from things that effectively work as a stand in for race because of historical

forms of discrimination. So I'll give you an example of that. Car insurance is more expensive in certain zip codes than others, and. Studies have shown that that is a racially discriminatory impact because of historical practices of redlining of making it near impossible for more African American people in the US to be able to buy homes in certain districts. And so that's what I mean by a proxy variable that sneaks in as a result of, you know, wider historical patterns of discrimination. Does that make sense. Yeah.

Question from the audience: I have a quick question that I don't imagine you have the answer to but I'm going to ask anyway. So I have a train of thought first, if that's okay?

Sarah Myers West: Sure.

Question from the audience: So, there's a person in (inaudible). Some people in this room know him and he has this really cool notion of social security, and its idea that there's security of systems and that presupposes that data has integrity and should be saved and information shouldn't be tempered, but socioeconomic security would prioritize the individual and the harm that could be caused to them. And so he was able to look at social media companies and how their business models which require an attention economy as a result of disinformation. I now see

how there's, you know, a business model problem with AI companies because they're trying to go to IPO, get whatever funding to get bought out right away. And it’s been pointed out to me that they’re actually just trying to create tools tools of utility that are just really useful. How do we incentivize the considerations of what you're talking about when they're just trying to go and make as much money as fast as possible. How do we change data scientists who are just curious or just driven by problems, like, you know, the two big issues, do you think there is a way to change these incentives (paraphrase).

Sarah Myers West: I mean, regulation is one one possible way to change the incentives. I mean, I'm, I'm wary of any approaches that sort of take the standpoint of like, it's just evil engineers that are out to create harm which is not what you're saying, but I think that that's often what people gravitate towards, is that like these people are trying to do bad stuff and we have to stop them from doing bad stuff. And these are really, really complex social phenomenon that are entangled with very deep and long lasting historical patterns of discrimination. So I think I'm addressing these challenges, probably needs to start there. Honestly, and also to take those dynamics into account in the ways in which they deploy. So, you know, if we know that there are the data that police are producing on communities is tied

in with dirty policing tactics historically, then don't feed that data directly into, you know predictive policing systems without, you know, acknowledging that that data is going to then perpetuate those practices, and that's kind of the purpose of the report that my colleague, a law review article, that my colleague Rashida Richardson wrote. And don't leave decisions to these systems or position these systems as if they're going to be making decisions in ways that are more objective and less biased, which I think is both, how they tend to be adopted. And that's also how they tend to be sold. So a lot of the hiring systems are positioned as less biased, more objective fix to diversity problems when really what they are a more efficient way of going through resumes, without, without addressing the problems. So, that totally did not answer your question. I know, yeah I mean that is that is the big one that's that's kind of driving all of the work.

Question from the audience: (inaudible)

Sarah Myers West: For those who couldn't hear he's talking about explainable AI and that's sort of, and I guarantee I'm gonna get this wrong because it's not an area that I'm expert in by any means, but it's these efforts to render algorithmic systems, more transparent and more understandable or interpretable so that you can identify problems before they get deployed in practice, and it's certainly an active area I

can't really comment too much on I do have colleagues aat AI Now that are looking at explainable AI. And I know that in the bias space, we do see this move. particularly by companies, to be really sort of fairness toolkits that purports to be able to, kind of plug and play models where you should be able to sort of like fix bias by like tweaking variables in solution. And those, I think are flawed for all of the reasons that I described because you have proxy variables that pop up because of the if you look from system design to deployment there's all kinds of places where discrimination can emerge that are not just statistical. But does that at least partially answer your question?

Question from the audience (paraphrase): Can you describe what goes into building the reports you guys build with discrimination, what does that process look like?

Sarah Myers West: Sure, right so, I started--the objective of the report was to try and understand the landscape of gender, race and power. And so I started by purposefully trying to scope it really widely across disciplines. So I ran lots of searches on different kinds of research databases. I looked at syllabi, I snowball sampled, so when I started with the seven studies I was looking at who they were citing, then looking at who they were citing. To start to fill out this really robust

body of data. I did everything that I could to try and find like little holes and plug in gaps (inaudible) To get into, like, the methods I then used thematic coding in my notes to identify what are the themes across, across the studies and used that to build out the report. That was specific to this it's not necessarily the same as other reports. I know that AI No--annual reports are more of a collective effort. We all have different things to bring to the table and things we build together to understand these issues.

Question from the audience: Instead of the diversification of data sets which you said is not always effective, how much of a role do you think the diversification of people sitting at the table could play in (inaudible).

Sarah Myers West: Really good question. Um, I think it's a really significant aspect. I think that it matters tremendously what people are empowered to do and what kind of workplaces they're working in. One of the problems with the pipeline approach is that we don't fix like a toxic or exclusionary workplace, you're kind of like feeding more diverse people into places where they're going to then hide their ideas, where they're then going to be discounted, where they are going to be paid less, where they are not promoted. I mentioned under (inaudible), That's a phenomenon that we see across the industry where women and people of color

might come in with a certain set of skills and are hired in job categories that are lower than those of others. And that has very long downstream effects about being able to be motivated and so what they experienced in the space really matters. Very much as well. And it's also not only about identity categories, it's about how different kinds of skills are valued. So, like, there's there's sometimes a privileging of technical expertise over other forms of expertise and having people from social sciences and humanities backgrounds can be really valuable in developing systems because of being able to bring lenses, to ask questions about what products are being built and what the long term implications of those products are so that matters a lot as well.

Question from the audience: Do you see a culture of accountability emerging in these tech companies, or do you think political intervention will kind of always be needed to ensure we get to that point?

Sarah Myers West: I mean, I think we are seeing a lot of instances where we have workers, organizing and companies making concessions. So, an example of that was the end to forced arbitration at both Google and Microsoft, which was a, you know, there was a pre existing policy that if you had a discrimination or sexual harassment claim that you would have to go into arbitration with the company. You

couldn’t sue them you couldn’t go into a class action lawsuit against the company, because those policies no longer exist. People who experienced harassment or discrimination can sue, they can actually get meaningful action and that action is going to be undertaken in the full public eye. So we see some moves there in the inside for sure.

Question from the audience: (inaudible)

Sarah Myers West: Yeah, no, it's a really it's a really critical question. I know, much less. One of the studies, the one about the 18% of the presenters at AI conferences for women, that study does look at diversity by country. And now, there are some methodological questions about how countries identify how juries identify across country because they are attributing gender on the basis of name which is not ideal methods. So, there's a little bit of data that I've come across not a ton. One thing that I have been looking at is the deployment of different kinds of systems in different countries. But I don't have as much information. It’s also partly because, to speak to your, your methods question this is often a really opaque pace. And so we're having to be really creative about how we go about studying them so trying to look at marketing materials, patents are one place that I found to be pretty

rich for trying to find out information about like what companies are doing and how to be creative (inaudible).

Alex Ketchum: Okay, well thank you everyone for coming.

-_____

Thank you for listening to this episode of our podcast. Transcripts for every episode are available at our website, disruptingdisruptions.com. This episode is also available in video format with captions on the series' youtube channel, with videos also embedded at disruptingdisruptions.com.

The Feminist and Accessible Publishing, Communications, and Technologies Practices Speaker and Workshop Series was founded and organized by Dr. Alex Ketchum of McGill University.

This series was made possible thanks to two Connections Grants from the Social Sciences and Humanities Research Council of Canada, the Institute for Gender, Sexuality, and Feminist Studies of McGill University, as well as McGill's Department of History and Classical Studies, Media@McGill, the William Dawson Fund, the Moving Image Research Laboratory, the Sustainability Projects Fund, the Dean of Arts Development Fund of McGill, the McGill Writing Centre, and Digital Initiatives of the McGill University Library. Further support comes from Concordia University's Milieux Institute for Art, Culture, and Technology; the Institute for Indigenous Futures; Machine Agencies; the Algorithmic Media Observatory; the Intersectionality Research Hub; the Black Feminist Futures Working Group; and Cinema Politica. From the Universite de Montreal (UdeM) we have received funding from the research institute MILA. Additional support comes from RéQEF (Réseau québécois en études féministes), MUTEK IMG Festival, Element AI, and L'Euguélionne: Montreal's Feminist Bookstore.

—

The theme music is from Lobo Loco’sOld Dancehouse - Long on the free music archive and InspectorJ’s "Bell, Counter, A.wav" of Freesound.org

Thank you to the series’ research assistants Thai Judiesch and Astrid Mohr for their assistance in producing the podcast.

Please subscribe for future episodes.

sudo apt install audacity