Jim O'Neill 0:09

Today's article is titled, an AI run the world needs to better reflect people of color. It's written by Angela Benton. The article is featured on wired. com, and you can find the link in the show notes. Now, here's the article.

Jim O'Neill 0:25

The AI market is expected to grow to well over $100 billion by 2025. We're just a stone's throw away from a voice activated, facially recognized, algorithm driven life. But for a rapidly growing segment of the population, AI can be more triggering than innovative. Much of the data being used to train machine learning algorithms, which power the AI movement doesn't take ethnicity or race into consideration to a lay person or someone disconnected from many of the day to day plates of people of color. This may seem inconsequential, or even race baiting. After all algorithms don't need to understand the user's ethnicity to make accurate recommendations and assumptions. That's the beauty of technology right? However, the more intertwined our lives become with AI, the more biases could bloom, some of which could result in life or death. Before AI exacerbates inequities throughout society, we must include and protect minority data today. errors from incomplete AI training data already affect people of color. For one facial recognition software has a history of Miss identify black citizens. disclosure. I am an investor in a facial recognition company that is champion not selling its data to authorities. Last year, the ACLU ran a test with Amazon's recognition software in which congressional headshots were matched against the database of mug shots. 40% of those misidentified were people of color, but they comprise only 20% of Congress recognition remains in use within some police departments. Amazon has also partnered with 400 police forces across the country. Which will use Amazon's camera doorbell product ring, whose facial recognition software is still in development to form a new fangled type of neighborhood watch. Also within the American criminal justice system, as a 2016 Pro publica investigation discovered, software used to identify future violent criminal threats ran on an algorithm that was correct only 20% of the time. Black defendants in particular, were pegged to be a 77% higher risk for committing future crimes than reality proved. Health care, which increasingly uses algorithms to determine diagnoses and treatments is also problematic. Nearly 40% of Americans identify as being non white, but 80 to 90% of participants in most clinical trials are white. This can be a huge issue for illnesses that disproportionately plague minority communities like diabetes or heart and respiratory diseases. In 2015, only 1.9% of respiratory disease studies included any minorities while I was going through breast cancer treatments, many of The procedures and therapies my doctors recommended were derived from studies that were predominantly comprised of white female patients. It was also extremely hard for me to get a referral for a mammogram when I was diagnosed at 34. Even though black women are typically diagnosed with breast cancer younger than white women are the recommended age to even get a mammogram is 40. Again, from data that disproportionately included white women, Dr. Joy woollen weenie and MIT computer scientist and advocate for ethical and inclusive technology says this coded gaze is a reflection of the priorities, the preferences and also sometimes the prejudices of those who have the power to shape technology. Meanwhile, a black and brown diaspora of data is quickly multiplying. In the US, people of color are projected to outnumber non Hispanic white citizens by 2045. Around 50% of the world's population growth between now and 2015 is expected to come from Africa. According to the Pew Research Center, a greater proportion of black and Hispanic adults use it Instagram, Twitter,

WhatsApp, Snapchat and YouTube than Caucasians. Facebook owns three of the top six most used social media platforms by people of color. That's an incredible amount of power. pair that with an estimated 37% annual growth rate of a penetration into business and racial bias will become an even more daunting challenge at the hands of our machines. As you will know, a hurry says those who control the data control the future, the AI that's being developed today will serve as the baseline for how AI will be built in the future. Data ownership is essential. It's not just our human right today, but also the key to our future rights when we need their own their data nor have a say in how it's used. We leave these decisions in the hands of a select few. algorithm with auditing is another promising solution. We can define what it means to develop and train algorithms ethically by reviewing the training data set diversity and biases of those who developed the algorithm. But this baseline way of thinking has barely permeated day to day business practices. And even if these reviews are implemented, we have to be mindful of the power structures that will employ them. Much of the power that exists in our drastically changing world comes from free market innovators who are accountable primarily to their shareholders, not their constituents. Some even have the capacity to make or break governments. Take for instance, Cambridge analytic is a parent company SEO groups work in election campaigns in Trinidad and Tobago, where digital attacks and data micro targeting campaigns were developed to sway votes. External credentialing institutions that set ethical standards law, the Bar Association for lawyers could be transformative. They're also necessary. Coding inclusive it into algorithms is a challenge when most developer teams are made up of paltry percentages of women and of people of color. But we must augment initial training data sets to reflect the actual demographic makeup of our society and allow consumers to opt in to share their data so that they are accounted for we must Design Technology that doesn't inadvertently oppressed those who have already been oppressed.

Jim O'Neill 6:07

This article was written by Angela Benton, the founder and CEO of stream Linux, which uses data science to measure what people are watching and listening to across streaming platforms. She is on twitter at a Benton. Now let's discuss.

Heather O'Neill 6:23

Hey, everyone. I'm Heather O'Neill. And I'm Jim O'Neill. And now we're going to talk about the article you just heard. And I run world needs to better reflect people of color, which was written by Angela Benton. So Jim, you just read that article for us. And thanks so much for doing that. There was a lot of really interesting information and sort of unsurprising for me, but what was your biggest sort of impression or takeaway from that article? Having just read it?

Jim O'Neill 6:49

Yeah. So I agree with you that the the a lot of the data presented was like, I didn't necessarily know the exact statistics, but it wasn't that surprising. You know, the Whole, like the fact that white supremacy dominates everything and influences all the data that are being collected. There were a couple of

surprises in there for me, like, for some reason, I had never really thought about the fact that the medical community is so affected by white supremacy and that it that it goes back further even then. It's not even necessarily about the data itself and algorithms. But just like with the received wisdom, about like, what you should do to deal with a certain medical issue is driven by studies that are have been conducted with primarily white people. So that was a little bit of a surprising thing.

Heather O'Neill 7:35

Well, that's really interesting because it's that was less surprising to me because of the history of studies only on men influencing how women receive health care or how products are built or like even the fact that a car is built for someone who's five, nine or taller, which is the average height for a man, which means that most women writing in cars, the sequels aren't really safe for them in the same way that they are for men. And it's all recently that they started testing safety against crash dummies that are smaller in size, have different shapes of bodies, all these other factors of what humanity actually looks like. And so for me, I guess that wasn't surprising at all. Yeah, that

Heather O'Neill 8:15

makes sense. You being a white dude. Exactly. have to think about it, which is the point.

Heather O'Neill 8:21

Yeah. What I also thought was really interesting is it harkens back to me about the the advance of color technology when it comes to photographs and once upon a time when you would photograph somebody in color. If they weren't a white person, just the color tones didn't show up right. And what's most I don't know if shocking is the right word. But like frustrating about that whole thing is the reason it got better was not because I people developing that technology to print color photos, were became more aware of the fact that like they were excluded A large part of the population, but because furniture makers wanted the Browns of their furniture and their wood pieces to show up more visually distinctly. And so this feels like an extension of that. And just another way in which mostly white communities are sort of not caring and not prioritizing people of color. And I think it's such an important part of the conversation because once something goes into an algorithm, there's a tendency to feel like, Oh, it's a machine so it can't be biased. And that's one of the sort of like the biggest lies that you know, we tell ourselves in technology is like, numbers don't lie. And you know, I talked about this a lot when I give talks on quantitative data numbers do lie. numbers can be faked and forged. You can clean up the data and decide certain things are outliers that drastically influence the results you get, and the conclusions that you You can draw, and also any data that you collect contains the biases of the society that they're collected from. And so unless you actively account for that, or you just reinforce the bias, right, it doesn't. Yeah, it doesn't make things better. It just makes bias now formulaic.

Jim O'Neill 10:17

Yeah, totally. The other thing that, to me was good to see not necessarily surprising, but was the the focus on how we are becoming, as I say, a majority minority society in many ways, and that non white people are going to outnumber white people. Both in general, and especially on all of these social media platforms that you know, I I didn't really necessarily realize that usage was, you know, it sort of going in that direction as well. And it made me think about how, as progressives, we sometimes get upset at the idea of the consolidating minority power in politics and how you know, the folks in power Sort of making changes to the rules so that they stay in power, even though they're no longer the majority. And it made me think about how the way that technology is set up. We're doing the same thing for the future, to consolidate the way we think about things in terms that are based on white people, even though we're not going to be the majority anymore.

Heather O'Neill 11:22

Yeah, and that's such a huge problem on so many levels, right? Because there's so much of a like, reactionary movement of trying for a large swath of life, people to stay in power. And so I think if you're listening to this podcast, you probably don't support white supremacy. And you do want to see it ended. And I would encourage you, especially if you're white, to go into your organizations, and especially if they're mostly white, if you work on something that's machine learning or AI driven, to ask some hard questions about what are we doing to account for the bias in society? What are we doing to ensure that our data doesn't just reflect the white popular What are we doing to make sure that we're not reinforcing racial stereotypes and biases that have been passed down from generation to generation? I'm specifically reminded of, I forget which state is using it. But at least one state has created a sentencing algorithm so that judges can't pass down by sentences anymore. But the data that it's based on is biased based on it's based on past judgments that have been handed down, which we know are biased, because the whole point of stopping judges from passing down sentences to reduce bias, but if you use pass judgments, then it's going to be biased because of judgments or bias. But it didn't occur to anybody. And this is actively being used. And so it's now it's now it's literally a machine based sentencing system that is fully biased. But again, because people have this disconnect, an algorithm seems like such a scary word or like this big nebulous thing. People think oh, Well, it must be, you know, error free. It's data doesn't lie. But data does lie. And if you feel like algorithm is a scary word and you're listening, just know that algorithm is just a set of rules that somebody made up that now the machine always follows. And the rules can be complex. But when you get dressed in the morning, you're using the algorithm and your brain of how to pick out clothes based on all the factors that you factor through when you choose your outfit. Like, are you going out that day? Or are you staying home? Do you have an important presentation? Are you doing a photoshoot? Are you going to the beach, all those things make up the rules of how you choose what to wear. But we would never call it an algorithm. But that's really what an algorithm is. It's teaching the computer, all those things. So the computer would be able to make dressing decision for you if you explain all the different factors that go into how you decide. And so when you're thinking about algorithms, just remember it's a set of rules that somebody wrote. And so if those rules don't account for Whatever biases are in the data that it's working off of, or whatever biases the person writing it has, it's going to be biased to.

Jim O'Neill 14:08

Yeah. So I'm quite interested in the two sort of branches of takeaways that are in the article in terms of data control and ownership, and then also data auditing, or sort of auditing our biases. And I am reminded by that of a few projects that I sort of am aware of that are in those fields that are sort of emerging today. But it is interesting to think about how we can all sort of get more involved in that in both of these realms. So for example, data control. I know that a lot of people in organizations are thinking about how might it be possible to envision a world where everybody controls their own data and parses it out or gives it or sells it to companies themselves, so that they control who sees it and music who uses For what and why, as opposed to just, you know, everybody's data being harvested unceremoniously by Facebook and the other companies that do this.

Heather O'Neill 15:09

That's an interesting concept. And I've definitely seen that discussion happening. But I think that also presupposes that everybody has an interest in actively controlling their data. Whereas I found that most people just want technology to work and they don't want their data to be anywhere. They don't want to be sort of tracked and gathered upon and, and targeted in the data way that the online world sort of works right now. But they don't have an interest in actively like managing preferences for their data. And so I like the idea of data control, but I think it has to be more of a nuanced discussion than it is currently where we just say all people can choose what they want. It's like when somebody says, Oh, well, we build the privacy settings. So you can do anything you want at any level. Nobody, most people don't want that. Yeah. And so what it really needs to be is Like a stopping where we need to stop these large companies that are, you know, run by white tech bros and sometimes tech ladies, who are just reinforcing white supremacy all over the place and not thinking about it and not caring because they don't have to. So we just need to prevent them from having access to this data or being allowed to use it. Like I think that's a large part of it, too, is giving that control back not in the saying, hey, guess what, now you can set up all your settings for how your data is used, and instead just saying, hey, guess what, we're not actually just going to track you anymore, and we won't have them. And so there's a difference in, in what that control really looks like. And I think it's less about giving individuals control though, I think that's a nice and important thing to have than taking control away from the people who've just sort of taken it. They didn't technically have no reason to not do it.

Jim O'Neill 16:53

Right. Exactly. And that's a great point that that sort of a two pronged, you know, idea that it would be great to empower the people who want to get into the weeds and you know, sell their own data or protect their own data and control it actively. But that means that doesn't mean that we shouldn't stop harming and stop tracking the people who just want to live. Yeah. So the other part of this then is sort of the data auditing or algorithmic auditing, as she says in the article that has a lot of interesting branches. To me. It brought to mind a story that I saw recently where Craig Newmark, the founder of Craigslist, partnered up with Consumer Reports and by partnered up with I mean, gave a boatload of money to and they are now going to be starting to review companies and products in terms of the Privacy Practices that they employ in their products so that if they're harming people, due to the way that they use data, or misuse data, those things will be flagged and there can be sort of a an established standard

for who does this well and who does it poorly. Of course, that makes me think also when you have a An auditing structure or an auditing, you know, system in place, who's auditing the auditors or, you know, like, what are the biases inherent in the auditing process itself? So it's, you know, it's it's biases all the way down.

Heather O'Neill 18:11

Yeah. And actually, I think this brings up a good point and actually a market gap in terms of roles that we should be hiring for. And as data science becomes such a popular, and like, sort of top job going into the future, because we are all swimming in data, we don't know what to do with or how to parse, there needs to be an equal sort of job that rises up that's like data auditing, and who controls the data? Who audits the data? What does that look like? And what are the standards? And I think that for people working in data or with algorithms or machine learning, who are the leaders in these community who should be talking about it, or who should be setting those standards, and how do we get some consistency across the board? You know, it's not impossible, but it's it hasn't even started right. I think Angela talks about that in the article. Just the idea of auditing an algorithm for biases is not permeating day to day business practices. And so how do we go from here to there? And I think it involves a lot of speaking up, especially on the part of white people who even if you're not involved in designing the algorithm, you're not doing the algorithm building or the math or the machine learning, you're working on the product that uses that. You have an opportunity and sort of almost a responsibility to say something to ask about the biases and to say, Hey, we need a way to do better and well, I can't do it. Let's find someone who can.

Jim O'Neill 19:37

Yeah, definitely. And in the, in the realm of auditing, and, you know, credentialing maybe or in setting standards, that's another nice you don't have to be making a product necessarily to have a role and to be able to speak up if you have a role in auditing or setting standards. You know, as soon as Angela in the article brought up the Bar Association, I was reminded have conversations that we've had about how when you involve credentials or certifications or things like that in professional life, you know, that can be a barrier to entry to people who don't know the right people to get the credentials in the first place and that kind of thing, or don't even money. Yeah, yeah, exactly. But so if you are involved in the idea of developing standards and things like that, especially again, if you are white, and you have institutional power, you know, use some of that to make sure that the team that you're on includes people from all walks of life and people who are going to be disproportionately impacted by the decisions that you make.

Heather O'Neill 20:32

Yeah, and I think that's a really good sort of final point to bring up, which is, if you're white, and you're in a mostly white space, talking about machine learning, and data and algorithms or working on something that uses those things. It's also your responsibility to say, Hey, we need more people of color in the

room. We need more people from groups who aren't represented here in the room because we're missing voices were missing perspectives, and we don't know what we don't know. And not just a Like we we work in user experience a lot. And so we do a lot of research, and not just as research participants, but as actual, like experts, consultants and internal team members. And so there's a larger conversation about how you build your team. And are you setting it up for the broadest level of success for the broadest number of people in the world?

Jim O'Neill 21:19

Yeah, so in, you know, a UX context or in a product context or anything like that, you know, often we want to make sure that we're asking questions like, how can we make sure that we're doing no harm with the things that we're building? But we often don't know the answer to that question, unless the right people are in the room unless we're including people who are really going to understand the ways in which that harm can be done.

Heather O'Neill 21:41

Absolutely. As always, Jim, it's wonderful to talk to you. And thank you everyone who's been listening. We really appreciate you. And if you have any questions or thoughts, you can leave them in the comments below or feel free to shoot us an email. This is Heather O'Neill, and Jim O'Neill from Tech thinking allowed Big thanks to this week's author for sharing the article. And thank you also to our producer Melanie Scroggins. To get details about anything we referenced in this episode, or to recommend an article for a future episode, visit tech thinking aloud.com. We'll see you in the next episode.

Transcribed by https://otter.ai