Live at AMIA 2022

Inclusive Informatics: Reducing Bias in Our AI Systems

Dr. Leyla Warsame (Host): Welcome to another edition of For Your Informatics, a podcast where we explore the limitless world of medical informatics. Created and led by the Women in AMIA, we offer insights into career paths, leadership, and education. Thanks for joining us as we highlight lives to inspire greatness inclusion and diversity in the field of informatics. Hello and welcome to the For Your Informatics Podcast, where we explore the limitless world of biomedical informatics. My name is Dr. Leyla Warsame and I'll be your host.

This special live recording at the 2022 AMIA Annual Symposium is focused on inclusive informatics and the quest for reducing bias in our AI system. Our podcast is being sponsored by the Women in AMIA Leadership. Thank you for your sponsorship and thank you for joining us. We're pleased to welcome our distinguished guests. Dr. Kevin Wiley, Dr. Kenrick Cato, Dr. Tina Hernandez-Boussard, and Mx. Adrianne Pichon. Thank you so much for joining us. Let's start off with brief introductions from our guests. Kevin, let's start with you. Tell us about yourself and your journey into the field of informatics.

Dr. Kevin Wiley: Sure. So good afternoon or morning to the millions out there listening. Right. So I'm Kevin Wiley. I'm an assistant professor at the Medical University of South Carolina. Prior to that, I was a PhD student at Indiana University, where my major focus was health policy. Minor focus was population health informatics. And so that is how and that, in addition to being a T 15 biomedical informatics and Data Science Fellow, helped sort of build my interest in biomedical informatics, specifically population health informatics. And so currently, my research is largely based on examining electronic health record data quality in chronic disease care, and where there's a large use of digital health care tools like telehealth.

Dr. Kenrick Cato: Thank you. Hi, everyone. My name is Kenrick. I am currently an assistant professor at Columbia University, appointed both in the School of Nursing and the Emergency Department. My clinical background is in emergency nursing and an oncology nursing. And in about two months, I'll be going to the University of Pennsylvania as a full professor at the School of Nursing and also appointed in Children's Hospital Philadelphia. I also have a background. I worked for a while as a database administrator and electronic health record analyst in It departments, both in hospitals and also in managed care. And my program research really focuses on applying data science methods to electronic health record data to support clinical decision making.

Dr. Tina Hernandez-Boussard: Hi. Tina Hernandez-Boussard. I am an associate professor at the School of Medicine at Stanford University. I also have appointments in biomedical data science, surgery and by courtesy and epidemiology and population health. I have had not a very straight trajectory into the field of biomedical informatics. Being a first generation into this realm of education, et cetera, mentorship and sponsorship has always been a huge aspect of my work and has really guided the focus of the work that I have done over the past. I would say ten or 15 years. My team has focused on how we can use very diverse sources of data to improve healthcare delivery, patient outcomes, and really improve the evidence that's needed to ensure that everyone has the right treatment and the right resources they need to move forward. We do a lot of application of novel methods to data. We use a combination of data, and it's all under the realm of improving the diagnosis, treatment, and patient outcomes. So happy to be here today. Thank you.

Mx. Adrienne Pichon: Hi, everybody. I am Adrian Pichon, a PhD student at Columbia University Department of Biomedical Informatics. I have also not had a straight path to informatics. My background is in public health, and I have this long-standing interest in sexuality and reproductive health and also chronic illness. Some of my earlier research was on menstrual trackers and apps for sexual health education, and my doctoral research focuses on patient generated data for endometriosis care and management. I am also a participant in the HL Seven Patient Empowerment group focusing on patient contributed data. And finally, I was a co-founder of the justice informatics group in my department with Dr. Oliver Bear-Don't-Walk and Harry Reyes.

Host: Wonderful. I am excited to welcome these guests, and I am actually inspired by the depth and the spectrum of experiences that they each will bring to this conversation. Let's start off with the first question. How does each of you define AI bias? And what was the canary in the coal mine that created this awareness? And finally, why is it an important issue for you?

Dr. Wiley: AI bias is, I think, the situation where artificial intelligence or machine learning tools, either by the data used, data that's trained for these purposes, either the rules and methods used in any applications by humans, create or replicate, reinforce unfair or unequal treatment for certain patient populations, specifically in healthcare. And I got involved with exploring, I guess, in depth AI and ML bias as a fellow at the National Committee for Quality Assurance, where we examined sort of on behalf of some health plans, what can be done in that space for enrolled populations. And I learned a lot starting back with a paper by Ziad Obermeyer that examined an algorithm that allocated resources based on healthcare consumption among races. So it's a good paper in nature, and that sort of really started to grow my interest in this space.

Dr. Cato: I fully agree with Kevin's definition of AI bias. I think I would just add two other kind of caveats that I think about a lot in my work. One is when individuals train models with a certain data set, and then they apply those models to data sets that aren't appropriate because the population of sample was not represented in the trained data set. And so that's one example. I think the other AI bias that I think a lot about is when individuals train models and they don't account for the bias that Kevin mentioned. That might reflect bias in the data set that's reflected or they don't really think about the bias that they are entering into the modeling process. So those are the other two kind of caveats to AI bias that I think are really important to think about. For me, actually, the canary in the cole mine came from me when I was working clinically, where I saw that on the units and it was specifically people who their primary language is not English. As a nurse, I saw that there were certain clinicians that weren't going into those patients room as often because they felt not comfortable communicating with those patients. And so because I actually was working as a programmer in the It department while I was working clinically, I knew that there was going to be bias in those data because there was less interaction with the patients, there were less observations being taken, the notes were going to look different. So it was very clear to me at that point that there was going to be biased in those data and then everything downstream from that was going to be biased, especially any clinical decision support that was attached to those data.

Dr. Hernandez-Boussard: Great. And so I think we're really getting some great definitions of AI bias, and I just want to build on that a little bit. And I think that AI bias, a lot of times we focus on the algorithms that are biased and the data that's biased. But I think it even starts before then. I think it starts with even the questions we're asking, right? And so a lot of times we're asking questions that might be able to be reframed to have a more fairer outcome. A lot of times we don't think about the societal harm or population level harm of these algorithms that were developed. So for example, do we think that the algorithms and the output of the algorithms is going to be fair across all populations? And so I think when we think about AI bias, I think there's many sectors of AI bias that happen before we even get to the algorithm. So for example, what type of data sources do we use or do we have to use to ask these questions? What types of aggregation of the data do we do? A lot of times you've seen this before. We aggregate race into black and white or Hispanic, non-Hispanic. And then we have this other category that is literally other, that is everything that doesn't have enough information to actually run and train models on it. And these decisions that we make and even thinking about the questions, thinking about the model, I think is really where AI bias comes. And in these decisions we make early on in our investigations, it makes bias run through the entire system. I also think that when we think about AI bias, there's two consequences that can happen. We know about allocation harm. So we have an AI system that will, for example, give biased allocation for certain services or outcomes or treatments. You can think about a risk score that is biased and therefore somebody might not have the equivalent of having an equal risk score to have resources, but there's also reinforcement bias. So when we're developing these systems, we're generating information. And that information we generated from these algorithms is also biased. So it's a cycle that we just keep building more and more information that's bias that's come from these algorithms.

Mx. Pichon: So for me, there wasn't really a canary in the coal mine. We've always been looking at bias and descriptions of bias. I think before we did a good job characterizing disparities and outcomes. Now we're doing a good job characterizing bias in AI. But I really hope that we can take this to the next level and think about how do we start mitigating these biases? So, going along with that, the communities that I really engaged with around queer sexuality and menstruation and chronic pain are communities that are already stigmatized and they deal with these messy embodied experiences that we're trying to capture and represent in data and then use these for computational purposes. And that's really hard when you have these really messy experiences that the contradictions are something that's really natural in human real menstruation and real lives. And so really thinking about the purpose of some of these technologies, why they were created and what were they created for? And how do they use the data and algorithms? For what purpose? Are they trying to redefine what a successful period looks like? Or are they trying to help people explore and understand their own bodies and generate knowledge that can help them to self-manage and self-care and things like that? And so, I noticed that the technologies that are being produced are really flattening these experiences, and the purpose is not for enriching these experiences. And those are things that I think are important to address.

Host: Wonderful insights. Can you please give us a brief overview of the work you engage in in reducing AI bias? Let's start with you, Adrian.

Mx. Pichon: Sure. So some of the work that I do in this area is about trying to rethink menstrual trackers and how do we design for irregularity as the norm and reflect the real life circumstances of individuals and how do we design for the varied bodies and identities and goals without predescribing what these things should be. That's really one of the earliest projects that I worked on and now moving on to chronic pain. It's really about how to help use these data and not flatten them, but help them to navigate power dynamics with their providers, how to empower users with their own data and the knowledge of their own bodies and illnesses and help them to work in partnership with their providers and themselves independently rather than having tension in those circumstances. So how can we leverage data to help people be in control and be in power in their own health care. Yeah, I'll stop there.

Dr. Hernandez-Boussard: So we do a lot of work in my lab thinking about how do we mitigate these biases? And there's a few things that we do that I think are really important when thinking about this. The first is team diversity. I think that you really need to think about the people on your team across all levels of professionalism, of education, of diversity, because the more diverse our team is, the more we can think about the questions we're asking and the potential societal effects. When you have diversity on your team, people can take it home, and how would this affect my community, how would this be engaged in my community? So we get a broader view of that.

The second piece, I think, is really in transparency. When you do a literature search at a lot of the algorithms that have come through for AI, there's really a lack of information on where the training data is coming from. What are the aggregations they've done for different variables? What types of models did they try or did they not try that failed or didn't fail? And so, really, having transparency across the entire process of the AI lifecycle, I think is really important. And finally, I think we fall short of doing the appropriate analysis to understand if our algorithms are biased. So we spend a lot of time thinking about not only the performance of the algorithms we're developing, but how reliable are they across populations, what are the different aspects of fairness that we can take out? Now, I think that all data is biased in some sense, but if we're transparent about that and really think about the populations we're applying it to, then we can use it in a responsible way. So we try and look at all those different aspects when we're creating our systems in AI.

Dr. Cato: So I actually just thought of something that I should have mentioned when you asked me the question of how I think of AI bias, that Dr. Hernandez Bersar just touched on. So I just want to be clear. For me, I think it's important to point out that bias is actually an important part of the work that I do. And it might be an important part of some of my colleagues in that it's really hard to do prediction without variability and bias.

I think it is important to kind of point that out. I bring that up because in the work that I do in the inpatient setting and also in the outpatient setting, I look at different levels of surveillance, especially nursing surveillance, of patients, to predict deterioration. And my team's work has shown that when a nurse is concerned about a patient, they start their levels of surveillance changes. And you can look at features in the HR to detect that and use that to predict deterioration. So there is a bias there that we harness to do prediction, which is you could define as positive bias. So for us, we fully recognize that those bias and surveillance also when we capture them, we are capturing there's a definite bias in our data of people that deteriorate and levels of surveillance.

So, for example, in our data, and this is data from New York City and Boston for non-white patients that end up deteriorating. And when I say deteriorating in the inpatient space, I mean things like mortality, unexpected transfer of the ICU, cardiac event, rapid response call, things like that, sepsis, bloodstream infection. For people who have those outcomes, there's a much higher level of surveillance for white patients and non white patients. And so in our work, it was very important for us to make sure that our prediction was bias, so that our predictive models weren't just detecting white patients potential for white patients deterioration more than non-white patients, because the actual data that we were getting was biased and was reflecting that bias in that clinical workflow. That's how we thought about that major bias. And so that works. We're starting work in our team. So our predictive models sit on top of software for clinicians and they see that prediction. We're starting work in our team where we're giving the predictions to the patients and getting feedback from them and asking them if they feel that their treatment is occurring in a biased way and feeding that back into our modeling. And that obviously takes a lot of science, both qualitative and quantitative, to capture that and understand that. In addition to that, we've also started thinking about how do you provide clinical decision support to a clinician to say that they're acting in a biased way. And that's a whole kind of program of research because there's a lot in the clinical decision support science that we have to figure out how to provide clinical decision support to actually get a clinician to change their behavior, especially around bias. That's the kind of work that we're doing.

Dr. Wiley: So again, a lot of my work is focused on assessing the quality of electronic health record data. And as the panel mentioned, there's a lot of biases inherent in data and data produced by humans, obviously. And so most of my work is again assessing the extent to which data are, for example, timely concordant between data sources like EHR data set and patient clinical notes before they're used to develop or train pipe, various pipe. AI or natural language processing pipelines and then some other data quality dimensions, including, like completeness, et cetera, are important for certain demographic groups, specifically in chronic disease care like type two diabetes. And so understanding the extent to which these data are complete, for example, prior to using them as a training set and ensuring also that, as the panel mentioned, they represent the populations under assessment are a lot of my focus, which are continually important. And then I think in addition to that is bringing attention to AI bias in populations where it's not necessarily discussed, and that is health plans. And so I think as part of bringing attention to AI and machine learning, biases is sort of sharing ways in which AI and ML biases can be detected and mitigated. And so that's where a lot of my work has fallen.

Host: Thank you. As with all things healthcare, I take it there's no silver bullet to fix AI bias. So how can we assure that bias is reduced in both older and newer AI systems? Let's start with you, Kevin.

Dr. Wiley: Yeah, so some folks out of Chicago created I don't know if folks in this room are familiar with the AI Bias Playbook. And I think one of the first things you can do is obviously inventory the models that you're using both past and I think existing to ensure that if you did perpetuate past harms, there are approaches for restorative means to maybe mitigate, if possible, past harms, and then existing models that may be perpetuating harms presently. And so the AI bias playbook does that. It also ensures for assessing the models that you have enumerated, testing them for various types of bias, be it ethical or statistical, which there may be overlapped. And I think following some sort of framework and I think Dr. Hernandez Boussard has one as well. I think following a framework that has been established and validated is an important approach.

Dr. Cato: Thank you. That's a tough question. I fully agree with the previous panelists, his points. There are also, I think, I dare say, policy levers that can be pulled. And I'm trepidatious about that because policy sometimes ends up being very blunt and there are unintended consequences that are not always considered. But I do think it's important to have policy behind something as important as this to try to make sure that bias is not proliferated. In AI and machine learning, I've worked in operational settings, so I know that, for example, in the electronic health record, there is decades of code base that is biased. And I shudder to think about figuring out how to rip that out, but I think it's necessary to be done. In addition to that, I would say, and I haven't done the research, but just anecdotally there's probably a majority of algorithms that are not necessarily AI based, maybe rule based or other that we use in clinical practice that are biased because individuals bias was not something that was considered when they were developed. And we've done comparisons with the work, with our work and some other early warning scores where we've detected bias in those scores, racial and ethnic bias, and some gender bias that are widely used. So I think it would be a heavy lift, but I think it's necessary for us to address those.

Dr. Hernandez Boussard: Yeah, and I could agree more with the previous comments. I think for me, there's two, and I'm going to push a little bit, because I know we don't like to say we need more policy, but really we need more policy. I think there's two aspects of this. One is that we don't really know how to monitor these systems once they get deployed into a healthcare system. So we don't have any standards to how to look at the societal effect, how to look at the effect across communities, how do we look at the effect once they're actually deployed in a care setting? Because most of these models we develop are just theoretical models that we do with retrospective data, and we don't know really how biased they are once they are implemented into the care setting. And so I think there's a big opportunity for the community to really think about what are the standards that we want to have in place before an algorithm can be moved to point of care. Do we just base this on performance alone? Because that's normally what we see when these algorithms come through, is they had an AUC of zero 85 or zero nine. And we don't really think about what are the standard metrics that we need to look at to ensure that they are fair and unbiased.

And then in addition, once they're implemented at point of care, how often do we monitor these? Do we need to look at these every six months? Every year? Do we need to retrain the data? What different populations do we need to look across? And so there's really a gap in standards and in policies to really make sure that these algorithms aren't biased. And then I'll even go a higher level at the regulatory level, when algorithms are actually going through the FDA, there are again, no standards for how generalizable they are across populations, how reliable they are across populations. So I think really what needs to happen is we need to agree as a community on what is the minimum set of standards we want these algorithms to go through without killing innovation.

And then also, what are the regulatory standards. And I think they probably are two different things. And then from a hospital system, how do you want to ensure that you're not harming patients that come through? So I think we have a lot of work to do, but I'll stop there for now.

Mx. Pichon: Yeah, those are all really wonderful points. So when I've been thinking about how to address bias in my own work, I really have been relying on these thought leaders in the field, like my fellow panelists. So I really appreciate all the wisdom that you all bring to the field. So I want to think about how we address bias in designing systems for patient facing tools. And the first thing is this idea of the data that flattens representations of illness and experience and that disproportionately harms the most marginal users, and it smooths out the nuance in what they are going through and can render them invisible. And so thinking about how do we address data collection involving domains beyond just what's strictly clinical, how do we address data definitions, how do we use and represent data? And something that's come up with the people who use the self tracking app for endometriosis is the importance of narrative and stories and using this to help guide how they tell their experiences of health. And then how do we then use those stories but then compute them for these tools, these AI tools that have a lot of potential for really supporting the needs of users, but at the same time needs to embrace the complexity and the contradictions and the nuance.

So I think that's the first thing as I move into my dissertation work, I'm thinking about the design of systems and really focusing on participatory design to leverage the lived experiences of these people and how they can be designed into the tools upfront. Human centered AI is something that has become really important to think about. How to leverage the benefit of computational systems and the intuition of users and the agency of users and how do we facilitate control for them and allow for input and changing the behavior of intelligent systems to suit the needs of individuals and to rely on their own intuitions and human intelligence.

As one of my panelists is talking about before, the transparency of these systems is really critical. And thinking about the end users as patients in addition to those as clinicians, it's important across the board. And I think there's some open questions there that I'm really hoping to address in explainability in data and methods and things like that.

The last thing that I am thinking about with how do we address bias in systems through my own work is really the responsibility. And going up to the regulatory and policy level, I think that is really important. I don't have a lot of experience there. And so those are things that I think how do we build the competencies for folks who that isn't their focus? Because when we think about AI tools, there is so much potential for harms, there's over promises there's inflating the power of these models and they have the potential to revise norms at a societal level. And that's really risky. And I think something that we need to as researchers and designers take that responsibility very seriously. And thinking about the purpose of these tools, what are we really designing them for? Is it meeting the needs of users? Is it helping improving health and quality of life? Or is it for the purpose of this broader system and those other endeavors that may not actually support patient care? Yeah, that's it.

Host: Thank you, Adrian. So I'll piggyback off your comment. So you guys are all in specialized fields. How can the normal person clinician, data entry person get involved in reducing AI bias?

Mx. Pichon: Sure. So I will start. I mentioned in my introduction the justice informatics group that we started at Columbia, DBMI. And I think that this is done more for improving my understanding of AI bias and ethics and justice than really anything else. We have created this community that really started as a grassroots effort to come and bring social and community support. And that's really where it started, was kind of peer mentors. But then beyond that, it really started as a community building effort. And we kept meeting and we kept engaging and we kept going deeper beyond our personal interests and started then connecting it to scholarship and research. And how do we connect justice to informatics and how do we leverage informatics for justice? And so I think creating these groups with interested, like minded individuals, when we gather, we have diverse experiences, backgrounds, expertise, and I find such value in engaging with clinicians and the partnership between, especially the school of nursing at Columbia as well as DBMI. So I think that bringing different types of people together for conversations and having very casual discourse where you can ask questions and have conversations and hold yourself accountable in a way that's not in the spotlight and that is not risking harms to others. And then at the same time, we have in turn become a resource and leaders. We held a workshop in AMIA on Saturday and so I think kind of encouraging the early career researchers out there to become leaders and to step into this leadership role to come to the table and to figure out how to work in partnership with the other experts in the field and to become experts ourselves.

Dr. Hernandez-Boussard: Great. And so I could not agree more with Adrian. I think for me, the three main things are Team Science, mentorship and sponsorship. And so team science is just a key, as Adrian was just saying, it's a key to bring people together and have these diverse views. You mentioned that we have very, not siloed, but very targeted research areas. But I cannot do my research without the clinician, without the nurse, without all of these other people that come and bring a different perspective. I don't know, I as a non-clinician, I don't know the questions that they face at the front line. I need to make sure that the questions we ask are relevant for the caretakers, for the patients, for the hospital systems. So I think team science is really such an important, integral part of how we think about bias.

The other piece is mentorship. I think mentorship plays a very big role in allowing people to explore and make these outside connections that we might not think about ourselves. So providing the appropriate mentorship for our students and our trainees to really explore these areas, these new avenues of addressing bias and mitigating bias, et cetera. And then the third piece is sponsorship. And in sponsorship, I think that's really where we're promoting people who are working in this area. We're making. Sure that they have avenues to disseminate their information, to talk about it. As Adrian was saying, we want to be sure they have platforms to speak and voice their opinions and get feedback in a way that they feel comfortable in sharing information and gathering more information. And the other piece I will say is what we don't talk about a lot of times is having the patient involved in these team science and aspects of team diversity, but also in mentorship and sponsorship.

So I didn't ask a lot recently when we talk about pulling for labeling bias for our AI algorithms, and I keep getting asked, do you have patients labeling the data? I'm like, oh, gosh, such an important question. And we really need to think about how we can really bring on more diverse people to get a broader view. So couldn’t agree more with what was said, and I just hope that that can broaden the conversation.

Dr. Cato: Yeah. And just to add to that, I fully agree with Dr. Hernandez-Boussard and Adrian and their comments. There's some other things. So this might be a little bit in line. I do a lot of work with the EHR and I get in trouble for saying this, but I feel like the EHR needs to be blown up and for us to start over in the sense that we tend in the clinical space, operational space, to design the workflow and the components of the HR for clinicians. But it's really to serve the patients and to help in their care and to make them healthier. And so we really need to have patients involved in the design of the EHR and how data gets in there. I think that would go a long way. And the reason I say we need to blow up the EHR is because that's the only way that's going to happen, I think.

In addition, I really believe in I've thought a lot about this term, data democratization. And one of the things that's really important in that concept, and I think it's really is for the people who enter the data, understand why those data are being entered and the ramifications of the way that those data will be used. And that requires a certain amount of knowledge about how machine learning works, uncertainty, data quality, Dr. Wiley works on. And we're not talking about PhD level understanding, but really high-level understanding of that information, because it is, I think, as Adrian had mentioned, that these AIs that we're using have really serious ramifications on people's lives. And so that's something that needs to be conveyed to the individuals that are entering those data, in addition to all of the user centered design that's been mentioned. So I think those are some of the things that I think about.

Dr. Wiley: I wanted to quickly pick up on a point that I think is also related to the question that was just asked. And Adrian mentioned invisible patients and patients can be seen as invisible in the EHR. And more importantly, there are patients who are not identifiable in the EHR as well who may not see any benefit from use of an AI or machine learning model or any other model for that matter. And I think there's an assumption that if you are in an EHR that you are simply sicker than patients who are not. I think that's referred to as informed presence bias. And it's something to consider in a lot of this work.

But I think it's important to reiterate, not only to multidisciplinary teams who might conduct this work, that we're all biased. And acknowledging that is, I think, one of the first steps to understanding how we might move forward with AI bias. I think ensuring command line checks when you're building your own models is important. And there are a range of ethical and statistical bias checks at the command line that can be used. I'm happy to share a few of them developed by Microsoft. I think there's a bunch of open-source ones as well. And I think that would help in analyzing and presenting any data to documenters, clinical documenters, inputting data on behalf of patients, to show that if they don't think they're biased, you could demonstrate that they're objectively biased. And again, reiterating the point that multidisciplinary teams that include clinicians, social workers, social scientists, et cetera, are important in discussing data relevance, data mismatch methods. Is this an appropriate framing of the issue of question is an important thing to do, especially if is it an equitable framing of a question? And then lastly, is the model specified to the data that it's being trained on? And so I think looking at these things and discussing them amongst a team who might be performing them early and often is a strategy that could be deployed.

Dr. Cato: Dr. Warsame can add one? So, my post COVID brain forgot one thing. The other thing. I think Adrian's probably heard me say this around the halls of Columbia a lot, and it's not an original thought, but I think it's also important. I really feel strongly in the education around ethics and ethical thinking and ethical dilemmas. And it's something that we need to incorporate into the training of individuals early in their career, so that when they are starting to learn how to do things like build models, they have that foundation. And they're thinking about the ethical ramifications, about the computational things they're doing or the qualitative things they're doing, or you can fill in the blank. So I think that's really important as well.

Host: Thank you all to all of our guests for your time and for sharing your insights with us. It's been exciting to hear so many things. The fact that AI bias doesn't start at the algorithm level, but even before, when you're asking the initial question that working in AI should reflect team diversity and therefore the community that it impacts and that bias is not all bad. Dr. Cato did explain how you can honor its bias in patient surveillance to do prediction modeling. And how do all of us get involved? We can have casual conversations with, likeminded, peers. We can turn the EHR on its head, and instead of focusing on the clinician and make it all about the patient, and we can acknowledge our own bias. These are some of the things that we have learned from our panel. I have one more question for the panel before we leave. Do you have any resources that you can point people to when they want to learn more about reducing bias in AI systems? Let's start with Tina.

Dr. Hernandez-Boussard: Great. I think there are so many resources out there today, and there's all kinds of courses and workshops. Having AMIA, for example, and all of the focus on ethics and bias is a great place to get started, to get connected, and to really identify where you can find your niche in this topic. There's also several Coursera courses on bias, and so I just think there's a wealth of information out there. I would start with AMIA, because this is a great place to connect with the community.

Mx. Pichon: Sure, I'll go ahead and go next. I think that reading books that are available kind of for the mainstream public, has been really valuable to our reading group. Some that I would recommend would be Dr. Ruha Benjamin's *Race After Technology* and her new *Viral Justice*, which I am very much looking forward to getting into. Another one that we really liked in our student group was *Data Feminism* by Katherine Dignazio and Lauren Klein. And there's a lot of resources on YouTube, on podcasts like this one, and they're really valuable to listen to with your peers and discuss them and start building a community of practice.

Host: Dr. Cato.

Dr. Cato: I was just quickly trying to search my Amazon Order history because it's a really good book that I read, and I can't remember the title on it, and I'll find in a second, but I have to second what Adrian said. I think there are a lot of really good books to read, and I think that it's important for individuals to continue to educate themselves because the science and the thinking around bias and AI does evolve and does get enriched over time. So I think it is important to continue to educate yourself.

Dr. Wiley: So I think another book recommendation is *Weapons of Math Destruction* is a good one. Very good book by, I think, a mathematician. And there are also a range of algorithmic bias assessment tools that can be performed at the command line. And I'll list a couple off I made a list a while ago. "Deon Ethics checklist for data scientists", that's a command line tool. IBM has "AI Fairness 360". That's an open-source toolkit that helps examine, report, and mitigate discrimination and bias and machine learning models throughout its life cycle. "Audit AI", that's a GitHub repository that enables detection of demographic differences in the output of machine learning models and assessments. Fair Learn, another open source, community driven project. And then "Ethical OS", which is a risk mitigation manual that provides eight risk zones for anticipating unwelcome consequences of buys and models. So those are my suggestions. Happy to chat more about them.

Host: I think FYI Podcast Book Club has plenty of books to look up now, so we're going to open the floor up for questions.

Audience Guest: Hi, my name is Benjamin Collins. Thank you for your attention to this topic. I'm an MD postdoc at Vanderbilt University Medical Center studying the ethics, legal, and social issues of AI in health care. One of the focuses of the conversation has been on how to mitigate AI bias, particularly recognizing that it appears very early in development, but at the same time, without good regulation, those biases will appear in tools that are in use. So I have really two related questions. What should people in health care working with AI models know about AI bias? Is there anything you believe they can do outside of the technical process to mitigate bias, as well as the same question about the general public.

Dr. Hernandez-Boussard: So I can go ahead and take a stab at answering that. I think that everyone, certainly in our nation today, is aware of the systemic bias we have in the society. And bringing that awareness upfront to leadership, I think is a great way to identify these issues. Right. A lot of times we think, oh, not in our system, not our data, but as we've heard, every data set has some type of bias. Just bringing that and highlighting that and monitoring that, I think is really important to bring it up to the importance of leadership. Mitigation is hard, right? It is embedded in our healthcare system. So we're not going to throw out every algorithm that we currently have. I mean, it will be chaos. So how do we identify which ones we want to address? I think that's a great question. And just starting to look at the effect of these models across populations, I think is key. I mean, you can see the work that's happened with the risk factors for kidney failure, right? And having race as an integral risk factor in that we're starting to break that down. We're starting to remove that. But it's through awareness and conversations and having that out there and that knowledge out there that makes it come to the point where we have to change. I don't know. I'll pass it on.

Mx. Pichon: Yeah. So I think it has to be an ongoing thing where we question these systems. It can't be a one-time thing. It has to be at every interaction with an AI system, at every interaction when we're designing these systems to really question and critique them and to know that the purpose of that is to make them stronger rather than criticizing them just for the sake of doing so or the researcher, anything like that. I think the other thing is using human intuition and really relying on what we as humans do really good at and using that to help calibrate the trust with the system rather than seeking to maximize trust with the system.

Dr. Wiley: I think echoing a point Dr. Hernandez-Boussard made about getting an understanding, like, at the regulatory level of just simply what an algorithm is. I don't think it's specifically defined for broad use by the FDA. And they call them software as medical devices, I believe, and they leave most of the regulation to an international, like, voluntary organization. And so I think maybe advocating, like, in a professional association. I don't know. Do you all know of any, like, AMIA or any larger associations that can help thoughtfully create some sort of way through public policy committee or something to start talking about these things and the issues that you're facing at your own institution at Vanderbilt? And I'm sure others are facing similar issues as well.

Dr. Cato: I think the question is asked about also the education of individuals. I think this is difficult for some AI approaches, but I think the work that I do with clinicians Explainability is really important. And there's more science that needs to happen for specific types of algorithmic training. But explainability is really important to understand the bias in how the models are working and then surfacing that to individuals at the same time. Individuals will then need to really be trained to understand that. And most importantly, I think it's important for people to be trained on various levels to understand uncertainty. Because I think maybe it was Adrian said something about adding human intuition. Part of that is also explaining the limitations of AI and these models and really being clear about that almost to the level of the cigarette package warnings. And so those are things I think would be important.

Audience Guest: All right, nice to meet you guys. Karen Baker from Duke. I'm a urologist. So no one likes to be categorized, but we do have to aggregate data or else then we kind of face the curses of dimensionality. Right? So I guess on the flip side of that, somebody has to put that data in, and frequently that input is really far removed from where that grant money came from. So here are my questions. How do you know if you're balancing that tension correctly between having to flatten or categorize the data? And then how do you get the resources to where that data entry needs to happen?

Mx. Pichon: I can go ahead and start with that. So I think the main point that I would say is really thinking about what the data are being used for and really making them fit for purpose for the particular task. And so for endometriosis, maybe sex assigned at birth and gender are not something that is necessary to bring up in a particular context, but maybe organ specific terminology might be, whereas something in diabetes, the organ specific functionalities might not be as important. And so that's a circumstance where I would say having a more broad, sweeping categorization can be protective for some people for allowing them to be kind of anonymized in the data set, whereas when it's really specific and necessary for them to be highlighted based on some sort of category or characterization of them. In particular, I think that's when the really detailed categories are useful.

I also did have a question about how do we incentivize this to happen, the resources and so I have that same question for my other panelists.

Dr. Hernandez-Boussard: I can go ahead and give another in addition, building on what Adrian said perspective of that. And I think it comes down when we talk about data sources we use, right? They are regulated through IRB, through financial incentives and all these things. And a lot of times when we look at data use agreements, we have to combine categories, we have to aggregate data so that we ensure that patients aren't identifiable. But if you look at the IRB and these data's agreements, I mean, they have not been updated in a long time. They are not fit for the type of data and the questions we ask today. So again, I think there's a great opportunity to think about how do we need to change this so that we don't have to aggregate and combine into these other categories, into these mixed categories. In the 2020 census, mixed race was the fastest growing racial category that they documented. But we have no category for mixed race. We don't know what to do with mixed race when we're talking about our EHRs and how we deal with this clinical data.

So I think there's a good opportunity to really push our regulators on thinking about how do we develop these policies that still ensure the safety of the patient, but allow us to do the appropriate type of analyses we need to do, looking at these specific subpopulations. So I think there's a lot of, unfortunately, policy and regulatory guidance that we need, or aspects that we need to ensure that we can get the information and the data we need.

Dr. Cato: Yeah, I just want to quickly add to that, because I think race and the HR is a good example. I think there's certain scientific rigor that quite frankly, I think people who do data science informatics with data they've got in the past. And I'll give you an example with race and the HR, I know from my work and lots of data sets, I've seen there's probably somewhere between 20% to 30% missing race data in the HR. But you still see a proliferation of analysis of race and ethnicity where that is not really addressed. People don't really point that out. To me, it's sloppy science and it's creating bias. It's creating bias and it's just bad science. And so I totally agree that we need somebody else to come in and say you have to do this. There are certain things that we need to do because we obviously aren't going to do it ourselves as scientists, I do think that more regulation is needed.

All right, so I had a quick question to Dr. Hernandez-Boussard. You had mentioned in your team composition that you have multiple levels of education. That really made me, it was really interesting. Can you just talk a little bit about how you achieve that? Because it seems really hard to operationalize.

Dr. Hernandez-Boussard: So I'll give you a good example of what we're actually presenting at 1:45 today, is we're looking at labeling. So we're looking at labeling bias. And so what we have is we're looking at depression in cancer patients and we're looking through clinical notes and we're asking different people to label. Does this patient have indications of depression, yes or no? So we have our senior oncologist who did you know, we can get him to do maybe 200 notes max. We have our medical students who are doing maybe 1000 notes, and it's really interesting. And then we have some people who aren't non medical environment, and we're looking at the bias across labeling and the diversity of our labelers and diversity of our labelers from the senior oncologists all the way down to the med students to non. And they label very, very differently as where some of our medical students are unbiased in certain aspects that they're looking for in the notes. So they're really scanning through for key words, et cetera. Our senior oncologists are like they'll read it and they're like no, they have an opinion. They've already formed, I think, just reading into the initial pieces of those clinical notes. So that's one way that we're bringing on these different levels of education, I think, that are really important and we are seeing important differences that can actually change the outcome of our classification model up to 15% based on who's labeling the data.

Dr. Cato: Yeah, so I implore you to please publish on that so that when I do that and my study section won't ping me for that.

Dr. Hernandez-Boussard: Yeah, we're trying. We're trying.

Host: This is Dr. Leyla Warsame concluding this special live recording of For Your Informatics Podcast where we explore together the limitless world of medical informatics. Thank you all for joining us.

Outro: Thank you for joining us for this edition of For Your Informatics, where we explore the limitless world of medical informatics. A podcast Follow us on Twitter, Instagram and LinkedIn at @FYInformatics, and never miss an episode. We would love to hear from you. Let us know what you think about the show, ideas for future topics or guests, and other suggestions. Until next time.