After the Fact | Race and Research: What's Next?

Originally aired June 25, 2021

Total runtime: 00:27:26

TRANSCRIPT

Jeannette M. Wing, Avanessians director, the Data Science Institute at Columbia University: These machine-learned models could decide whether someone should get parole or not. Whose garbage gets picked up first, who gets the low-income housing. The very technology that we're inventing, the very technology that has such promise to help society, to benefit humanity, is showing that it can be biased.

Dan LeDuc, host: Machine learning, computer algorithms, artificial intelligence. That kind of technology will drive important, life-changing innovation in the coming years. But as Columbia University's Jeannette Wing just noted, the data that underlies that technology can contain historical biases—the past presenting problems for the future.

Welcome to "After the Fact." For the Pew Charitable Trusts, I'm Dan LeDuc, and today we conclude our series on race and research by looking forward. An adage of computer science is, "garbage in, garbage out." How can we ensure that the data that goes into the algorithms that make both big and small decisions—a consumer's credit score or a defendant's jail sentence—are based on unbiased data?

We'll start with the big picture. Lee Rainie at the Pew Research Center led a survey of technology innovators, developers, business and policy leaders, researchers, and activists, and more than 600 of them responded to this specific question: "By 2030, will most of the artificial intelligence systems being used by organizations of all sorts employ ethical principles focused primarily on the public good?"

Sixty-eight percent of those respondents said, "They didn't think so." That's our data point for this episode, and it underscores a wide concern among experts about the systems we are building to solve future problems. Here's Lee Rainie.

Lee Rainie, director, internet and technology research, the Pew Research Center: We asked two questions in mid-2020 that we are now reporting on. One, we asked, what's the new normal going to be like in 2025, after the pandemic has washed through the globe? And then we asked a

separate question about what will ethical AI, particularly AI design, artificial intelligence design, look like in 2030? Because there are so many people now who are talking about artificial intelligence applications in our everyday lives.

Dan LeDuc: In that latter report about ethical AI, it's fairly recent, what were some of the questions you asked? And how did your technology experts respond?

Lee Rainie: Well, we asked them what would be the prospect for ethical AI design, so how much would ethics be baked into artificial intelligence systems and processes by the year 2030. And more people expressed worry than expressed delight about what was going to happen. But what's so interesting is, even the people who were most downcast about the prospects for the future, have plenty of interesting things to say about what the upside might be at the same time that they have concerns. Same thing with the people who have some positive views about the future, they can well articulate some of the things that draw their concern between now and then. And, so, there were a couple of answers that really stood out in the broad themes of this report. The first of which was, that there was a big, sustained argument about what ethics means in the first place. There are different contexts in different cultures for what fair means. How do you privilege certain organizations over others? What does justice look like?

And, so, the starting point was we can't even agree on what an ethical framework might be. And that will be a problem, particularly as different societies apply these tools. Another main concern of many of the respondents here is that the central practitioners of AI systems are the big technology companies, like social media companies and other technology firms that are dominating the internet. And they control a lot of the data. They control harvesting of the data from us, and all of their users. And then they apply these algorithms and systems of learning to that data. And that gives them tremendous power over what's being studied, what's being predicted, what kind of outcomes that they want. In many cases, people were worried that the profit motive, for most, would be guiding their decisions, rather than things like public good or the fate of democracy, or things like that.

And, finally, they talked about the fact that artificial intelligence systems are already in place and operating. And, so, to the degree there might biases in those systems, they're already being baked into the culture, or baked into the way in which people use technology now. And they say the genie is out of the bottle. And, sometimes, it will be hard to do the forensics about what's gone wrong and get things right. And, sometimes, we won't ever understand exactly what some of the potentially harmful impacts will be, because they're now embedded in people's lives in ways that are just hard to unpack.

Music transition

Dan LeDuc: Before we dive into the deep end on how the machines are taking over, we turn to Jeannette Wing, the director of the Data Science Institute at Columbia University, to tell us more about what data science is. And how she and her colleagues are working to use "data for good" by identifying and addressing biases.

Jeannette Wing: I like to give a very short definition of data science. Data science is the study of extracting value from data. And the most important word in that definition is value. Because value is subject to the interpretation of the end user. So, value to a scientist is discovering new knowledge. When you think about all the data that astronomers collect through all the imaging and the satellites and the telescopes, discovering the next exoplanet might be the value the astronomer gains from data.

Data science is in our daily lives without our even realizing it. Many companies, of course, collect data about us, about people. And they use this data about people to make money, sell ads, sell us more products. Every time you go online to shop for groceries, for books, every time you watch a movie, data science is behind what those services suggest that you might want to buy—making recommendations. What you might want to next watch.

Data to a company is actually, potentially calculable. Value to a business is bottom line. And then data value to a policymaker might very well be informing through data what that mayor of the town should do in terms of zoning laws for new residential houses on a coast that's continually being flooded because of climate change. Value, really, is subject to the interpretation of the end user. The second important word in that definition is extracting, because it takes a lot of work to extract that value from the data.

Dan LeDuc: As we explore the notion of the influence of race or the lack of the influence of race and how it should be accommodated in research, how does that apply to data science?

Jeannette Wing: One of the most advanced data science methods today is called machine learning. And there are particular methods within this general area of machine learning that work best when you have a lot of data. So, the more data you feed the algorithm, the better the model that does the prediction or classification. In order to train—that's what we call it—in order to train the model, we feed it lots and lots of data. And, typically, all this data is historical data. So, if I am using machine learning to build a predictor on how someone might behave in the future, it will be using data from the past.

Now, the problem with all of this is that we know that historical data is biased. So, if we think about specific uses of these machine-learned models to make decisions about people, then we can really hit home the problem with potential racial bias in these machine-learned models. So, for instance, these automated decision systems, these machine-learned models, could help

decide whether someone should get parole or not. Whose garbage gets picked up first, who gets the low-income housing.

It's nice to think that these automated decision systems can help automate the process of making these decisions to actually counter human frailty and judgment because sometimes we get tired, we have a bad day, and we're not uniform and consistent in our own decision-making. To kind of decrease the variance across humans and their inconsistencies in judgment is the whole idea and hope of using machine-learning models.

So, the very technology that we're inventing, the very technology that has such promise to help society, to benefit humanity, is showing that it can be biased.

But if we can actually think about these issues of systemic racism, bias, and so on as we're inventing the technology, then hopefully we can prevent bad things from happening and certainly before any of this technology is deployed.

Dan LeDuc: You mentioned earlier that these automated decision systems that can help counter human frailty and judgment. Some of this new technology is being used to determine parole and probation. It used to be if someone was sentenced to a crime or commits a crime, gets convicted, there's a whole investigation that's done into their background. And they figure out, do they have a good support system? What's going on in their family life? What's their job? And then a judge has to sit down and sort of make the decision, how do I sentence this person? Now, these algorithms are sort of preparing all of that for the judge. The defense lawyers don't have a chance to sort of challenge the process because those algorithms are usually like proprietary. It seems like we're getting whole new questions of fairness in society.

Jeannette Wing: I want to talk about fairness specifically within the context of risk recidivism and parole and so on. Because it was actually an article that came out in a popular press that raised the attention that some of these automated decision systems used in the U.S. court system are biased. In one article, they claimed that these systems are biased against Blacks over Whites. And this caused quite an uproar in the academic community.

The company that deployed the system claimed that, in fact, their algorithm, or their model, I should say, is fair. And the article that was claiming that the system is not fair used a different notion of fairness. So, there were these two notions of fairness being used—one in one case to argue that the system is fair, and the other in the other case to argue that the system is not fair. And it turns out both notions of fairness are perfectly sound and perfectly reasonable. And, intuitively, they both make sense. But what the academic community pointed out is that these two notions of fairness are inconsistent. In other words, it's mathematically impossible to satisfy both notions of fairness at the same time.

Dan LeDuc: What were these two notions?

Jeannette Wing: So, one notion is called statistical parity or sometimes it's called demographic parity. And it just says if you have two populations, like Blacks and Whites, the rates of false positives should be equal for Blacks or Whites. That's very intuitive. It's very natural. And it's used commonly all over the place. The other notion is subtler. It takes into consideration the outcome of the system. In other words, did the defendant re-offend or not?

And what was interesting, just to make it concrete for the listeners, is that the article that was accusing the system of being unfair noted that the Blacks who did not re-offend were given higher risk scores than Whites who did not re-offend. So, they were wrong on the false positive rate for Blacks in that sense. But as a double whammy, Whites who did re-offend got a lower risk score than Blacks who did not. So, this was another reason there was so much uproar about this particular system. It's really surprising how actually a real-world problem came to spur the academic community, that community invented that technology, and now they're trying to solve the problem that they invented themselves.

Dan LeDuc: You know, the notion of introducing data into these sorts of situations inherently is trying to make it basic and fair, right? You rely on the numbers.

Jeannette Wing: That is the irony. There have been studies in the past to see how human judges are not consistent in their decision-making. You are not actually the same from hour to hour, from day to day. But it's worse than that, of course. It's not even ensuring some kind of consistency over an individual's judgments. It's ensuring consistency over many judges. So, you really want to flatten out the variance across different people, but it's even more than that. You want consistency over different jurisdictions. Let's just take one state. You don't want one county's court system to be harsher than another county. And then we know from watching TV shows that people find out who their judge is going to be, and they kind of can already predict whether that judge is going to be harsh on them, or not.

Dan LeDuc: And those technical questions also raise ethical ones. You've written about data ethics before; explain to us how that applies to your work and the technology community.

Jeannette Wing: When I think about data ethics, I do think about ethics in general. I also look to my colleagues in other professions, medicine, journalism, law, business. All of these professions teach their professionals, their students, ethics from day one. It's ingrained in a doctor that one makes ethical judgments in treating a patient. The technology community has not thought about ethics from day one. And what I am trying to promote through the Data Science Institute at Columbia, and part of my tagline "data for good," is the importance of data ethics.

We don't want our technology to be harmful to individuals or to be used in harmful ways. And so now we have to figure out, working with sociologists and lawyers and so on, we have to figure out how to remedy the situation.

Is there a way to look at the data and say, well, if you use this data, then, of course, you're going to get a biased outcome? Is there a way to de-bias the data? And what does that even mean? You might de-bias it along one feature and then add bias somewhere else. So, this is like cutting-edge research.

Dan LeDuc: In short, the coders, the experts, decide which variables matter and how algorithms use that data. More now from Lee Rainie.

Lee Rainie: If you can identify which variables have been undergirding the biases and prejudices of data, you throw them out. Or you readjust for the way that they are understood in the model, or to say don't do that anymore. There are ways to write code that essentially is, do the opposite of what this was originally doing. So, there's a sense that this is manipulable. It takes advantage of new skills that computers allow people to deploy that they didn't have when they were just calculating on abacuses and little hand calculators and things. So, there's a sense that this new suite of activities can be deployed in the furtherance of fairness and justice and righting wrongs and even identifying wrongs in the first place.

The biggest question, of course, is who gets to decide at the end, whether it's going to be companies that are looking for profits, or public-serving agencies or actors who are thinking about more than that. And there's definitely a sense that this figures into the larger debate about inequities heading into the future around digital divides, where the technically competent get to run far, far ahead of the folks who are not necessarily masters of technology.

Dan LeDuc: That gives us a sense of optimism, that there's a recognition of [the] problem, and looking forward. But, of course, to really benefit from AI right now, we still have to rely on the old stuff for a long time. What do the technology experts say about how they can take decades, maybe centuries worth of data and use it in an effective and fair way?

Lee Rainie: There are a couple of ways they talk about that. First, they talk about putting humans in the loop. So, as these systems are being designed, let's think about average users. And let's think about outlier users who might misuse this stuff. I know any number of artificial intelligence teams at the bigger corporations go through these wonderful red team exercises, where they get a request from a client, like, "Can you create me a system that does this?"

They go through an active brainstorming process to say, what is the worst that can happen? How can this thing that we think might turn out really well and maybe make us a lot of money, how can it be misused in the wrong hands? Or, how can it even, in and of itself, generate results that

are problematic and challenging to people? So, there are ways, at the dawn of the process now, that lots of this is being considered. The second is, that there is an ever more robust forensics community trying to understand what the black boxes do. Sometimes, they can't figure it out. And that's one of the big concerns these experts have. But a lot of times now, there's an intentional focus. We don't understand exactly if it's "garbage in." But we're going to make sure what happens at the back end is as fair and as unbiased as possible. And, so, there's that level of thinking about ethical AI design. And, so, I know there are ways to make the 2.0 version better than the 1.0 version.

And if anything, the computer science world has taught people that there's the alpha testing. Then there's beta testing. Then there's the launch. And then there are the patches. Or even entire rethinks that are done when the system doesn't actually perform ideally the way that it should or can be tweaked to do even better than it was originally designed to do.

So, there's the march of progress, both at the front end of the process and at the back end of the process, where some of the worst stuff can be caught before it creates too much damage.

Dan LeDuc: But I have to say, in reading the report, these computer guys make the point that, in many cases, these algorithms are being done in the business world. They're being done for a profit motive and that the profit motive weighs in and makes things happen fast or, as you said earlier, produce a good financial result for a company. That doesn't inherently include all this other stuff of checking in and making sure things are reflecting things accurately, or in a nondiscriminatory way.

Lee Rainie: Exactly. This is a prominent concern of the respondents to our survey that, much more so than in the public sector, or in the nonprofit sector, the for-profit sector is dominating the AI field and, in many cases, responding to governments or responding to organizations that would like AI applications applied to themselves. So, there is a lot of talk among these experts about how do you surround that system with accountability mechanisms, with measurement mechanisms, with, literally, rules of the road that could be applied to the systems, where they just don't go rogue. Or they don't escape logical boundaries when you're thinking about justice, when you're thinking about trying not to be malevolent, that we can clearly articulate this is not the way to go, or this seems a legitimate way to go.

So, there are ways in which putting ethical considerations into these systems is a growing insistence in the policy world. The other thing that's always striking in the technology world is there's a hacker community out there. There's an open source community that is bound and determined to hold the for-profit folks' feet to the fire, and to road test their systems and kick the tires, not only to see whether they work, but to see whether they are performing in an ethical and justifiable set of reasons.

So, there are some counterpressures out there on these systems. And, yet, the big concern gets back to the purpose of the series you've been holding this whole season on the podcast, that there's a fundamental sense that technologists are deeply implicated in all of the inequities that are emerging in our society. In the "New Normal" report that we put out, there's a sense that, to the degree that technology is now more embedded in people's lives, and people are more dependent on it, those who do well, those who can afford the best technology, those who have the most tech savvy, tech literacy, are just going to run rings around other people.

There's a sense that all kinds of inequities are going to be amped up as technology becomes even more centrally involved in people's lives. And, so, there's a big policy debate about how to head that off or how do you mitigate that once it becomes evident. And it's taking place at the same time [as] this gigantic conversation about the future of work itself. The jobs of the future are going to require a lot of tech savvy.

Dan LeDuc: Lee, this whole season, we've been looking at race and research as America becomes a much more diverse place and how the research world is being affected by these changes. How are those issues playing out in the conversations you have with the experts on technology around the world?

Lee Rainie: When we asked this question of our expert panel last year, it was just after the killing of George Floyd and the protests that ignited around the world. It was on everybody's mind. And a bunch of these technologists, essentially, made the case that artificial intelligence, and all the questions about the ethical design of it and the impact of it center on race and ethnicity. If you can't get past the biases and prejudices of our past using these new tools, then they're no good. First of all, there's a sense of worry that the application of these tools to the degree that it is built on data that arose during periods of bias and applications of unfair practices, the data itself might be teaching the new systems the wrong thing, or to embrace the wrong prejudices. So, that concern is top of mind for lots of these analysts.

Dan LeDuc: We've been talking a lot about all the bad stuff that could happen and the fact that discrimination could be carried on. But one of your experts, Esther Dyson, who is I believe in Switzerland, said that, "AI can reveal previously hidden patterns, the better for us all," she says. "And that a lot will depend on society's willingness to look at the truth and to act and make decisions accordingly." So, maybe the other way of looking at all this is that we do have an opportunity to see some new patterns and information about ourselves.

Lee Rainie: That is the most hopeful note that many of these respondents sound, that the best way to fight the problems that AI can create is to deploy smart, good AI against that bad stuff or at least to diagnose what the bad stuff might be and remedied by another application of the tool. There's a sense that AI will help expose AI's problems. And Esther is a friend who has been a

regular contributor to these reports. And one of the constant things that she keeps saying is that humans are still in the saddle. We can make the decisions about what to deploy or not. We can make the decisions about where we're going to orient the algorithms. We have the capacity in ourselves to point those forensic AI tools at the AI systems that might be creating problems. And, so, it's human judgment and human morals and ethics that very much can control how this unfolds. And her constant call to us, and particularly in the answer she gave here, is be your best, because humans, at their best, can do a lot of good things with the tools that we make for ourselves. That's the story of history and why shouldn't it be the story here.

Music transition

Dan LeDuc: And so that work continues. Thanks for joining us for our look at race and research. As always, you can learn more on this subject at our website, pewtrusts.org/afterthefact. And our work at "After the Fact" continues as we plan for our next season. You'll be hearing from us about that soon. In the meantime, we hope you'll subscribe wherever you listen to your podcasts. For the Pew Charitable Trusts, I'm Dan LeDuc, and this is "After the Fact."