



SuperDataScience

SDS PODCAST
EPISODE 1013:
WEAPONS OF MATH
DESTRUCTION, TEN
YEARS ON (WITH DR.
CATHY O'NEIL)



Jon Krohn: 00:00:00 What makes an algorithm terrifying? My guest today says it's not the complexity of the math. It's the secrecy, the unaccountability and the fact that you can't opt out. Welcome to Lucky episode number 1013 of the SuperDataScience Podcast. I'm your host, Jon Krohn, a decade after her mega bestseller Weapons of Math destruction sounded the alarm on algorithmic harm. Today's guest, Dr. Cathy O'Neil, Harvard Math PhD, former Wall Street Quant and host of the new AI Skeptics Podcast. She's busier than ever. Through her algorithmic auditing firm Orca and her nonprofit Ocean, she now provides the statistical evidence behind lawsuits against some of the world's biggest tech companies. In this episode, Cathy Puncture's AI Hype explains why AI won't so much replace workers as degrade them and lays out how all of us can demand accountability. I've waited to have this exceptional conversation for 10 years.

00:00:57 Enjoy. This episode of Super Data Science is made possible by Anthropic, Notion, Excel Data, and Gurobi. Dr. O'Neil, welcome to the Super Data Science Podcast. It's such an honor to have you on the show. How are you doing today?

Cathy O'Neil: 00:01:12 I'm great. Thanks for having me, Jon.

Jon Krohn: 00:01:14 Of course. Yeah. We've actually wanted to have you on the show for ages, but it's one of those things where I was like, Cathy O'Neil is so big, we'll have to have some kind of in. But somehow a month or so ago, I just summoned up the courage. I was like, "I'm going to find a way." And I think we got through to you through your agency, through your speaking bureau or something like that. Really appreciate it.

Cathy O'Neil: 00:01:34 Yeah. It's my pleasure. I like doing podcasts. I like having conversations.



Jon Krohn: 00:01:39 You have your own actually that you launched

Cathy O'Neil: 00:01:41 Recently.

Jon Krohn: 00:01:42 Yeah.

Cathy O'Neil: 00:01:42 I started the beginning of 2026 a podcast called AI Skeptics where we are skeptical about the hype, but also we're not anti. It's just like, how good is this really? How are the financials really? How much does this cost? What's the cost benefit analysis really? Because we think people are asking the wrong questions. So that's where we ask what we think are the better questions. I

Jon Krohn: 00:02:11 Guess a lot of the discussion today around AI hype is this idea that in the very near future, some people think months and then a lot of people seem to say years that we'll have full automation, at least of digital work. Do you think that's overhype? It seems to me like that's probably overhype, right?

Cathy O'Neil: 00:02:30 Full automation of what kind of work?

Jon Krohn: 00:02:32 Of digital work where basically anything that could be done by a remote worker will be... I mean, some people would consider that, I don't know, some kind of like AGI threshold where you could just have full trust of these digital employees that are AI systems working fully autonomously and as competently as an expert human in any kind of digital role so that it would cover legal work, software development, data science stuff that we do, writing books, just kind of everything.

Cathy O'Neil: 00:03:08 That is so beyond what I think is reasonable or plausible. One of my favorite little puzzles or maybe conundrums to ask people when they ask me questions like this is like, if you read an AI generated review of a restaurant that looked amazing, would you go there to eat? And the answer I think is absolutely not. AI reviews of food. It's like I believe that they can make my mouth water



theoretically. I mean, maybe someday in the future they will successfully convince me because they've stolen a bunch of reviews from people who actually are reviewers. They will successfully slap together something that will make my mouth water, but that doesn't mean that they know what they're talking about. The answer is no. I don't think that's going to happen. Also, I just like look at the stuff that's being generated now. Have you tried to read some of that stuff?

00:04:15 It's awful.

Jon Krohn: 00:04:17 Quite.

Cathy O'Neil: 00:04:17 Have you listened to AI music? Have you gotten AI emails? They're awful. So no, sorry.

Jon Krohn: 00:04:28 And there's something about it where it seems like it's relatively easy to tell as well because you kind of get these styles that are just wherever this kind of sameness.

Cathy O'Neil: 00:04:40 It's incredibly bland because it's almost like a smoothing out of any related conversations that have been found on the internet, including unfortunately now AI generated crap. So it's like smoothing out of things that were already smoothed out. So if you think through that, eventually it's just all going to be utterly flat. But I guess part of me is just like, it's either going to be really boring or it's going to be really annoying or a combination. If you think about customer service representatives, it's already been replaced by AI. Let's be honest. It's been a long time since we've actually been able to call customer service and get a person on the first go. And even all my tricks for getting to a person are failing because they really, really have just decided they're not doing that and they're getting away with it, but it's not because it works well, it's because they can and you're only calling because you're desperate.



00:05:36 So it's one of those situations where you're just going to be like, when I have to interact with AI, it hurts.

Jon Krohn: 00:05:41 For sure. Well, you have the podcast that's relatively new and I encourage listeners to go to it. What you're perhaps best known for is a book that you wrote a decade ago called Weapons of Math Destruction. And to give our audience a little bit of context on your background before we get into the book, you have a PhD in math from a little known institution. I think it's pronounced Harvard. Something like that. And then you held academic positions at MIT and Barnard College before becoming an analyst at DE Shaw, which is one of these Wall Street quant funds that is extremely famous and one of the most challenging ones to get into. But your experience on Wall Street and as a startup data scientist led you to actually join the Occupy Wall Street movement after witnessing how data actively targets the vulnerable and algorithms actively target the vulnerable.

00:06:34 And then from that experience, it seems like you wrote Weapons of Math Destruction, which became a mega bestseller exposing how opaque unregulated algorithms function as engines of economic inequality, surveillance capitalism, institutional distrust, labor control and consumer exploitation. And yeah, we're going to be talking about all that kind of stuff in this episode. But first of all, what about your bio did I not get right or was I roughly on the

Cathy O'Neil: 00:07:02 Market? Yeah. All of that is true. I guess I would say just to reorder it just a slight bit, after the financial crisis, I left the hedge fund for a risk firm. Then I was super disillusioned by how little anybody actually cared about risk because they were all being bailed out and especially the big banks. And that's when I joined Occupy. And it was back in the era, if you can remember way back when, Google still said don't be evil was their motto. And so when I left finance and joined as a data scientist of the



tech firms, I really was like, "Oh, thank God I'm not doing bad stuff anymore." And it was like by being educated, in fact, by my occupy colleagues, I ran a working group up at Columbia University for 10 years actually.

00:07:58 I was educated by my colleagues about how to think about power and who are the usual victims of bias and that kind of thing. And I started applying that to what I was doing in data science in ad tech. And so that's what I kind of like, wait a second, hold on. All this terrible stuff we've been talking about that happened in finance is now happening in big data, which is the marketing term du jour back then, now it's AI, but it's basically whatever it is that you're supposed to be intimidated by and trust somehow. So I wrote the book Weapons of Math Destruction, which did come out 10 years ago, which is amazing, but I conceived of it 14 years ago. It takes about four years to write a book. And I was pretty early and I was lucky to be early on that beat where I was just like, "Actually, this is terrible." Yeah.

00:08:54 And it was basically like a cult of arms, if you will, to the public. Because trust me, I tried to alert people in data science, "Oh my God, I've seen this before. I saw this happening in finance. Let's not let this happen here." And everybody was like, "Cathy, shush." And I was like, "Oh wait, that's the same reaction I got in finance. Shush, stop talking. Don't rock the boat. This is working for me. I'm going to get a great bonus. I'm going to buy a nice house. My kids are going to go to a fancy school." It was just really gross, but probably very normal.

Jon Krohn: 00:09:31 That is kind of the Wall Street perception of the external perception of Wall Street, I suppose, is that kind of anything that you can get away with, whether it's legal or not, if you can get away with it. Ideally it's legal, but maybe kind of in a gray area.



- Cathy O'Neil: 00:09:51 Yeah. Usually it is legal, but it's just exploitative. And I guess what I'm trying to remind you of is in the world of tech, this was actually not the culture. At least it wasn't acknowledged to be the culture. It wasn't rapacious capitalism back then. Silicon Valley was still a little bit like libertarian, information needs to be free. We're going to make the world a better place. I'm not saying it was anti-capitalist. It never was, but it wasn't how do we exploit people type stuff until it was. And now it really is. And now we know that. Nobody's surprised now to hear of that, but it was an evolution.
- Jon Krohn: 00:10:33 Yeah. And we'll get into more content on the book in a moment, but there's something that is maybe particularly interesting to our audience, which is a lot of hands-on data science, AI practitioners, which is that prior to Weapons of Math Destruction, because you talked there about how it takes about four years to write a book. In 2013, you had a book called Doing Data Science: Straight Talk from the Frontline. You co-authored that with Rachel Shutt or Shoot. And it's based on, I believe, a Columbia University Intro to Data Science course, which must have been at that time in 2013, an intro to data science course. It couldn't have been that many around.
- Cathy O'Neil: 00:11:13 We actually invented that course, me and Rachel, and then I live blogged it. And you can still find it on Math Babe, my blog at the time, which is still live. And then I turned it into a book. And it was really an examination, to be honest. This is how early on it was. What is data science? Is it just statistics? Is it just computer science? Does it deserve its own name? Because back then statisticians were like, "You guys are just stealing our techniques and our ideas." And computer science were like, "Oh, that's not really a thing." So we were just like, "Yeah, well, maybe it is a thing." It was a premise of the book was like, "If it's a thing, what is it?" And so we invited a bunch of active data scientists to come say what they did at work. So it was quite simple.



00:12:02 It was kind of a straightforward idea. But I think the book was... By the way, the other reason I wrote that book is because I had just gotten a book agent because I already had the idea of weapons. Oh,

Jon Krohn: 00:12:19 I see.

Cathy O'Neil: 00:12:19 I

Jon Krohn: 00:12:19 See.

Cathy O'Neil: 00:12:20 Yeah. And he was like, "Well, why should you be the one to write a book that takes down data science?" And I was like, "Because I'm a data scientist." He's like, "Well, okay." And I was like, "Well, okay. I'm going to write a book called Doing Data Science." So it's like a calling card. I'm the one who wrote the book. That was the idea. So it was proof of my credentialing. There's a lot of credentialing going on if you want to write a book. Obviously that makes sense. I mean, I'm not a writer. I mean, I didn't start out as a writer. I started out as a mathematician and I made a big pivot as it were to be a writer. And my agent was just like, "Show me you can write. Show me that you have these credentials." So that was part of it.

00:13:04 But the other part of it is I'm literally curious, what is data science? What's the future of data science? And one of the things that we got right in that book was the ethical concerns right off the bat. Although most of it is really like decision trees, things like that.

Jon Krohn: 00:13:21 Yeah. We were a lot less vulnerable, I suppose, back then when these automated systems were so much simpler where random forest or support vector machine is kind of as complex as it gets. I mean, I suppose there's still a lot of room for exploitation there. But now, I mean, even 2016 is when weapons of math destruction came out and a ton has happened since then in terms of AI capability with large language models, deep fakes, short form video



being created by GenAI, surveillance pricing that has largely become dominant since 2016. Do you think we need like a weapons of math destruction sequel that is going to talk about all these kinds of high powered AI capabilities that make us even more vulnerable?

Cathy O'Neil: 00:14:12 I think the things that make algorithms terrifying aren't their complexity or their technological advancement. It's a combination of the secrecy, the unaccountability that nobody in particular is in charge of checking if it's right or in charge of mistakes. And the fact that people absolutely have to use them. They don't have choices. They could be flow charts. They could be linear regressions and often are logistic regressions. It don't have to be complicated for them to be terrifying and exploitative. So I mean, a lot of the algorithms that I wrote about in weapons that are still there and there's probably the most terrifying algorithms of all. I would argue there are new ones. I would say the AI chatbots that tell kids to kill themselves would rank up there as one of the most terrifying of all. So I'm not saying there's nothing new under the sun, but I'm saying like the ones that I wrote about are still awful and they haven't been addressed.

00:15:22 To answer your question though, could there be a sequel? I actually considered writing a sequel. One of the things that changed my mind is that everything is changing so quickly that it would, as I said, it takes four years to write a book. It would be completely outdated by the time it went to China to be published and then came over on a ship. It's 12 months later to be sold. And that's why I started my podcast. I literally started the AI skeptics podcast because I was like, yeah, it takes too long to write a book nowadays. You can't address these issues four years from now. But yeah, I guess I've always considered the complexity to be independent of the terrifyingness of an algorithm, of a system. Absolutely the most terrifying one of all for me was the one called recidivism risk algorithms that sometimes judges were using to sentence



people to longer in prison based on risk score of them someday getting rearrested, still being used, still sending people to prison for crimes they have not yet committed and might never commit.

00:16:35 And by the way, simple as pie. Sometimes they're just questionnaires and then points are added up. It's stupid how simple they are. But I still think they're unconstitutional. I don't understand how they're used.

00:16:48 It just doesn't make sense to me, but the power is in the algorithm and the power is what is terrifying. That

Jon Krohn: 00:16:55 Makes a lot of sense. Since 2016, it seems to me like there might be some places where we can now be using algorithms that previously would have been ineffective. So in chapter seven of your book, for example, you mentioned how productivity management models are optimized for efficiency and profitability, not for justice or the good of the team. And for a long time, things like the movement of hourly workers in warehouses, fast food chains were monitored to enable automation and other efficiency gains. But now with large language models, we can be doing things that we couldn't do before. So for example, Meta announced its model capability initiative, MCI, I think that was this year, just a few months ago, which is a plan to track every keystroke and mouse movement of employees to train its AI models suggesting that the kind of demeaning degrading of labor that has long affected blue collar workers on say factory floors or fast food chains is now affecting even the highest paid echelons of white collar work at places like Meta.

00:18:04 Yeah. There wasn't really a question there, but I feel like you might have a response.

Cathy O'Neil: 00:18:08 I especially like the way you ended that, which is like I interviewed truckers and teachers and people who had gone to prison or are being denied parole based on



recidivism risk algorithms. Let me just give you a little non-answer to your non-question. Those are the people I was talking to when I was writing the book and I was like, this is demeaning, this is humiliating, this is degrading. It's dehumanizing. Truckers are working for algorithms. They're working for surveillance systems. They're not working for themselves anymore. I used to, by the way, load trucks in high school. And I met a bunch of cowboy truckers back in the day in the '80s. They were just absolutely cowboys. I mean, they were also hopped up on all sorts of different kinds of drugs, but really interesting, very, very interesting kind men, mostly men and independent, really independent, like independent businessmen.

00:19:15 And it's very small business, usually just themselves. But that no longer exists, that model of the cowboy trucker. It's just not possible anymore. And that way of life just being removed and being replaced by something that is much, much less interesting and sexy and human, it was sad. But true. And the teachers that were being fired based on almost a random number generator, similar, teachers have been underappreciated, underpaid because back in the day we had free labor basically from women like Louise and Mary Alcott or Laura Ingalls Wilder. They were just like super smart women. They were only allowed to teach so they were underpaid. Anyway, the point being that I was working with people who are workers and they had a very strong notion even back in 2014 that when I was interviewing them that like, yeah, these algorithms are working against me. This is something that's happening to me.

00:20:26 This is happening to me. And then I remember I was asked to give a TED Talk the year after my book came out because it was pretty popular books as you say. So I went to the TED main stage and I gave a TED Talk and I talked to some audience members and they were so excited about big data. And they were just like, "Oh, this is



happening for me. This technology is happening for me. It's going to make me more productive. I'm going to have... " Elon Musk came the same time that I came. Even back then by the way, I just was like, "This guy is a jerk." And I'm holding myself back because I know you bleep people on this podcast. But people were just super into this idea that Neuralink, they're going to have an automatic connection to the internet and they'll be smarter.

00:21:19 They're going to be smarter and faster and more productive and things are happening for them. And I just was like, "This is the new divide." And that was 2017. I was like, "This is a cleavage in our society, which is only going to get bigger." That's what I meant when I said how big data increases inequality and threatens democracy. The cleavage is the people who think the systems are working for them versus the people that think these systems are working against them or it's happening to them, which is even more like sinister way of thinking. But to go back to your point, Jon, and sorry to ramble, but for me, what's happened with AI and the engineers whose keystrokes are being measured, blah, blah, blah. Which by the way, that happened to me in the hedge fund. All the keystrokes were being measured then too, just for spying issues.

00:22:15 But yeah, so white collar workers are getting what blue collar workers have had for many, many decades, which is a humiliation and degradation and dehumanization. And in some sense, I think that's a good thing. Not that it's a good thing, it's not, but it's good to have more people on the side of the workers. It's good to have solidarity to understand, yeah, this is real. This is happening. AI is stealing your ideas, stealing your work. You are no longer necessary because we've already stolen the stuff that we need. Not really true though. That's the good news. Going back to my skepticism about whether



digital work is really going to be replaced by AI. No. Some stuff will be, but not all of it.

- Jon Krohn: 00:23:07 In several episodes of your podcast, you talk about something called Taylorism in reference to this discussion. What is that?
- Cathy O'Neil: 00:23:14 Well, taylorism is this notion, and this is a really excellent book called Monopoly Capitalism. And I'm trying to remember the name of the author because I have terrible memories for names, but maybe we can look that up where he talks about Taylorism and Taylor. And basically it's the germ of what ended up becoming like management consulting. I don't know how much your listeners know about... They probably know about McKinsey. They've probably heard of management consultants. But did they know that back in the day, workers had artisan skills, like metal workers. There's a great chapter about metal workers, like people who would get orders for metal to be shaped in certain shapes and they would use all of their knowledge about, okay, how thick is this metal? What kind of blade do I need? How do I need to turn this while I'm cutting it? What kind of cutting implement do I need?
- 00:24:20 All these machines that were extremely complicated and they were experts. And their managers such as there were because there weren't that many managers back then were just like, "Oh, could you get that done? Do you need more metal?" They were kind of working for the artisans and Taylorism is flipping the script is basically being like, "Study the workers, write down all the information that they are using, turn it into a science that information is now owned by the management class. And the management class instead of having an artisan, is going to have a factory floor of artisans. It's going to have many, many machines that are doing different things. And at the end of the day, the metal workers are reduced to people who just feed sheets of metal into a machine that



has been programmed to end up with the result that has been ordered.

00:25:15 Now, you could argue that all of that is progress and it's much more efficient. And in a lot of ways it is progress and it is more efficient. But it also, the point of terrorism and the discussion of terrorism is that it means that the management class and the owners and the capitalists have all the information, all the power and can replace workers with less and less skilled people who are very replaceable and have no power.

Jon Krohn: 00:25:44 I think the book that you were describing is by Paul Sweezy and Paul Baran.

Cathy O'Neil: 00:25:52 Nope. Different one.

Jon Krohn: 00:25:54 Oh, okay. Monopoly Capital was the name?

Cathy O'Neil: 00:25:57 It's a book by Braverman. And I'm trying to remember his first name, but I want to say Alexander, but it's Braverman and I think it's called Monopoly Capitalism. Excellent book. Everyone should read it. And it really describes not just metal workers, but what happened when email was invented? What happened when the computer was invented? What happened to all those secretaries? I don't know if you ever watch Colombo like I do. I watch it once a year. I watch the entire series again because I never actually remember. I have this wonderful memory that fails to remember who kills who and why, but I just love Colombo so much. But I love all those scenes where he walks into a business, like a magazine office and there's like 40 secretaries in a room and each executive has at least one secretary, but maybe more than one secretary. You never see that anymore.

00:26:57 So what happened to those people? This book explains it.

Jon Krohn: 00:27:01 And so the book I found it, it's called Labor and Monopoly Capital. Sorry about that. The Degradation of Work in the



20th Century, Harry Braverman. But interestingly, it actually builds on the book that I first found called Monopoly Capital. So the first book, Paul Boran, Paul Sweezy, Monopoly Capital, and then this labor and monopoly capital book by Braverman. Apparently, I mean, I'm literally just reading the opening paragraph on Wikipedia. But according to that opening paragraph, they're somehow related. And the labor and monopoly capital book is obviously newer in 1974. Yeah.

- Cathy O'Neil: 00:27:39 And the labor and monopoly capital book, which I'm looking at on my shelf right now, so you'd think I would have been able to remember it, is itself like a neo-Marxist. So it relies on some of the notions that Marx himself came up with, which many of them were of his time and didn't really make sense in this world. So in some sense it was like, okay, let's extend this. And by the way, Jon, the book I wanted to write that I decided would take too long because it's four years is basically like, I want to take Braverman's book and just rewrite it for the world of AI. I want to extend it another 50 years. He wrote it in 75, by the way. It came out in 75. And in 75, he was talking about what was happening in 1925 with Taylor. And I wanted to be like, okay, 2025, what's happening now?
- 00:28:32 I still think that's a book that could be written and it could be written in a way that doesn't age as quickly as the world is. I mean, because a lot of these ideas are very evergreen. Workers being degraded, humiliated, dehumanized and replaced by themselves, but paid less and with fewer skills. We're not really being replaced by AI and that's not going to happen, but I do think we're going to be degraded.
- Jon Krohn: 00:29:04 I think my perspective on this podcast, the line that we generally toe is this techno optimism. But I do also definitely acknowledge there's lots of negatives. Well,



- Cathy O'Neil: 00:29:15 What is your techno-optimistic take on this? Let's have a discussion.
- Jon Krohn: 00:29:20 Yeah. I mean, it's tough to come directly out of these kinds of things that you're describing. I don't disagree with anything that you've said in the episode. And I definitely, I mean, I experience personally the way that... So we recently crossed over a thousand episodes and for episode 1001, the guy who founded this podcast and hosted it for the first four years, so the show's 10 years old for the first four years and a bit, the host was Carol Aramenko. He founded it. He still owns the podcast business with me, but I've been hosting it now for about six years. For episode 1001, Kiral interviewed me as kind of this role reversal where I'm the guest for the first time or actually I guess I was back when he was host before there was any discussion of me becoming host, there was an episode in the 300s where I was interviewed.
- 00:30:19 So 700 episodes later I come back on, I'm interviewed by him again. And the kind of headline thing that I think ended up being the YouTube thumbnail and the title of the episode on YouTube is how I spent 20 years getting a PhD focused on machine learning applications and then in industry in financial markets like you as well, then in digital advertising and in startups. And so had this kind of moat around my technical skills that I developed over all these years. And now it would be crazy for me and it would be crazy for a lot of people to not be using a tool like Claudecode to rapidly be able to test different machine learning ideas. And there's a very little point in me today kind of hand typing keystroke by keystroke my Python library imports and a training loop. I can do that if I enjoy it kind of for fun, but not really on the clock, if that makes sense.
- 00:31:23 And I can definitely say that a lot of the technical moat that I built up around myself for many years has disappeared.



- Cathy O'Neil: 00:31:32 And I'm going to be completely honest with you, I am kind of a techno optimist. I believe technology is absolutely amazing, but we have to be very careful about what we're optimistic about. I'm optimistic about the power of technology to do amazing things. I'm not optimistic about our choices, political choices really, geopolitical choices on how to treat people as this stuff is happening. I'm not a fan of China, but I heard, and maybe you know more than I do about this, Jon, so just tell me if I've got this wrong, but I've heard that the Chinese government has told people, do not replace workers with AI. And a few workers were replaced by AI sued and won. They got their jobs back. And I'm not saying we do that because it's very much not free markety, but it points to something which is that this is a choice.
- 00:32:37 We are choosing how to treat people and it's a political choice and it's a cultural choice and we don't have to make the choices that we are making. But unfortunately the people who are in charge are just like very immature selfish people who do not feel at all obligated to a social contract. And so the question for me is like, how do the workers who have solidarity hopefully by this point or soon, how do they press their power towards the very small cabal of tech executives who are making all the decisions? So at the end of the day, it's like a question of power and it's a question of like voice. It's a political question, but it's not really anti-technology. I mean, imagine like if we had really great, this is kind of the question of like, do we like democracy or do we want like kings?
- 00:33:41 And the answer is always like a really great king would be better because Congress is always mired in indecision and in fighting and stuff like that. There's all sorts of terrible inefficiencies by design really in our government. We'd rather have a really good king, but we don't have a good king. Or you can't depend on a king to be good. And right now what we have is like a very small group of



non-good kings in the world of technology. But if we had... So again, my point being is that it's almost neutral to the question of technology itself.

- Jon Krohn: 00:34:20 The argument that you're making there allows me to believe that there's comfortably a book, A Weapons of Math Destruction two. It's kind of interesting. I guess the title maybe doesn't even work as well now in the 2020s or 2030s because it really plays on the weapons of mass destruction thing, the WMD thing that was a big news item in the Bush era, which maybe some of my younger listeners are not even really like on top of.
- Cathy O'Neil: 00:34:46 Yeah, exactly. No, it's definitely an old pun.
- Jon Krohn: 00:34:50 A Weapons of Math Destruction two style book that updates the Taylorism concept from a century ago as you say. And by the way, I'll have links to articles on Taylorism as well as the Labor and Monopoly Capital book for our listeners for sure to follow up on that. In case people are curious right off the bat, the Taylorism comes from someone Frederick Winslow Taylor, who is, he's passed away more than a century ago. But yeah, exactly. The resources I'll point to, they talk a lot about the management style. And actually this is a bit of a tangent, but do you know Kurt Vonnegut's first novel? It's called Player Piano.
- Cathy O'Neil: 00:35:33 Don't know it.
- Jon Krohn: 00:35:34 So Kurt Vonnegut imagines a future, maybe like the 1980s, the 1990s, the distant future from the perspective of the 1950s where the kind of computing technology that was new in the 1950s, so using magnetic tape to store information and having punch cards. And so he basically, he has factories where it's just magnetic tape and punch cards that are running everything in the factories and you just have a few people with PhDs. So even executive



assistants in his imagining have PhDs because there's so few jobs available for people to have anymore.

- Cathy O'Neil: 00:36:14 I love the idea that executive assistants still existed. Yeah, it's like he saw data centers before anybody else. That's great. I mean, by the way, a lot of, I mean, Kurt Vonnegut is a positive example of this, but a lot of the problem I think, I mean, this is just me projecting, but a lot of the problem I think coming from the political decisions or the sort of moral obligations that aren't being felt by the tech overlords comes from the fact that they read a bunch of dystopian science fiction when they were young and they thought that was like cool.
- 00:36:56 They're like, "Let's make that happen. Cool. Terminator. Yeah, let's do that. " It's almost like they're just like they watched Star Trek or they watched Star Wars and they were on the side of the empire. They just didn't get it. They didn't understand the prime directive. They didn't understand the spirit of the prime directive. By the way, I have three sons and when my older two sons were little, I would make them watch Star Trek, the next generation with me. And then we went on to Voyager and Deep Space Nine. We went to all of them, even The Enterprise, which I love. I still love that, especially the theme music is excellent. And we were just for some reason last night around the table reminiscing about how wonderful the lessons learned there were about...
- 00:37:49 I remember because we were talking about how Riker was always like, you could count on him to like, if needed, have sex with an alien. If that was going to get us out of a bind, he'd be like, "Okay, I'll take one for the team." Sometimes he did it literally because he's like, "This is how we're going to save the crew." But the other times he just fell in love. And we were talking about this one planet where people weren't allowed to have a gender, but this one alien was like, "I secretly feel female." And he was like, "Let's do it. " And they went off to the holodeck or



wherever they do that kind of thing. Anyway, I'm circling back because I'm on a tangent of a tangent. But yeah, Kurt Vonnegut saw into the future as many science fiction writers do. And the question is, do they have enough signage in their imaginings, in their short stories or novels to say like, "This is not a good idea.

00:38:52 Don't do this.

Jon Krohn: 00:38:54 " Yeah. And I mean, that is the whole premise of player piano. And of course the player piano name, of course, comes from recording somebody playing a piano and then the piano just playing by itself. And the idea is that these factories are full of, like you were describing the metal workers where you find the best metal worker in the factory is the idea in player piano and in real life. And you use to his imagining in the 1950s magnetic tape and punch guards to remember exactly how the metal worker did what he or she was doing and have that be done perfectly. And so then there's this town of people. They have universal basic income for everyone. That's

Cathy O'Neil: 00:39:36 Where things break down.

Jon Krohn: 00:39:37 So there's all these people almost everybody's on UBI and they're not happy. They're very unhappy. They have no sense of value. They feel degraded and they come up with some schemes, which I don't want to ruin any plot points, but basically the people try to come up with a plan.

Cathy O'Neil: 00:39:57 Is it Luddite-esque?

Jon Krohn: 00:39:59 Exactly. A Luddite-esque without trying to spoil anything. But something that's interesting in the book is, so you talked about executive assistants because basically in the book, Kurt Vonnegut, he mostly fleshes out this idea of blue collar work being automated. But there's an interesting chapter, which I was hoping he would get into more later in the book, where the main character in the



book, he's a manager of one of the biggest factories and he is an outstanding backgammon player. And he can beat all the new recruits at backgammon. And so whenever new people join the company, there's this kind of ceremonial thing of him beating them at backgammon. But this group of young engineers who are new to the company create a backgammon machine, which of course is a huge thing that has to be rolled in. And it appears that it's going to beat this backgammon star, the main character of the book at backgammon.

00:41:03 But then somebody literally throws like a wrench into the machine or something to cause it to go haywire. And there's a line around that time about how it's like it's that guy, the main character, this engineer saying to his executive assistant, oh, how they're lucky that their work could never be automated, that only the blue collar work could be automated. And so I feel like he was kind of like touching on that with the backgammon thing, but it never gets fully fleshed out.

Cathy O'Neil: 00:41:32 He was still brilliant. I actually had the pleasure of seeing him speak at Cody's books in Berkeley when I was a college student there.

Jon Krohn: 00:41:40 That is amazing. He's my favorite fiction author. I think I've read all the novels. I haven't read all of his essays yet or all of his shorter form content yet, his short stories. But yeah, I think he's really funny. And one of the things, this is now we're on such a weird tangent here, just basically

Cathy O'Neil: 00:42:03 Talking about

Jon Krohn: 00:42:04 Kerfonigate.

Cathy O'Neil: 00:42:04 We'll just hang it.

Jon Krohn: 00:42:06 One of the things that I really like about him is that he doesn't really ever have bad guys.



- Cathy O'Neil: 00:42:12 That's a really important point. Someone actually asked me the other day, "If you had three bullets and you could kill three people and you're not going to get in trouble for it, who would they be? " And I was like, "Nobody. I'm not killing anybody. These people are all products of the system. They're products of this terrible shit system that we need to get better. We need to improve the system." I actually, having been at a hedge fund, having spent time around very rich people, billionaires, they're totally socialized to be absolute jerks. And look at me editing my words here.
- 00:42:51 Everybody tells them how smart they are. Everybody agrees with everything they say. Can you imagine being in that environment for longer than a week without becoming a flaming jerk? Well, it happens. It happens to every single one of them. They also stop caring about other people. They're megalomaniacs. And you've seen this. I'm sure we've all seen this. We've just seen this happen in front of our very eyes to the people who were just like one time Harvard undergrads with a techie nerd idea and then they become who we see today. It's the system. You don't need bad guys. And that's what's actually kind of scary about it. It's scarier that way. Bad guys are defeatable. And that's one of the reasons I'm not a superhero fan. I don't watch those superhero movies. It's just none of it works. It's not how bad shit really actually goes down is like where Superman could intervene and put an end to it.
- 00:43:50 Just not true.
- Jon Krohn: 00:43:52 You and I are on exactly the same page in terms of our perspective of how the world works as well as film and fiction novels. And then it actually ends up tying really nicely to weapons of math destruction. So we ended up being on long tangents, but we circled all the way back. Wow, look at us. Because it is basically this idea that we have just over time created systems, Vonnegut-esque



systems that are kind of not doing the vast majority of people a favor. So Vonnegut, a lot of it is about war, but a fair bit of it is justice at work and other kinds of things like that, that weapons of math destruction tackles head on. But in all of it, in the real world, in weapons of math destruction, in Vonnegut books, it is a system that kind of gradually evolves into what it is that is creating injustices and recapitulating injustices, having LLMs memorize injustices and regurgitate them.

00:44:53 Yeah. So I guess it's kind of nice that we kind of come full circle. Now is maybe a great time to talk about what you're doing to try to change the system a bit. So you have an algorithmic auditing company called ORCA, which stands for O'Neil Risk Consulting and Algorithmic Auditing. And then you also have a nonprofit OSHAN, which kind of recursively has ORCA as the first O in its - I

Cathy O'Neil: 00:45:18 Like that, right?

Jon Krohn: 00:45:20 Yeah, exactly. ORCA Collaborative Expert Assistance Network. So yeah, tell us about ORCA and Ocean and how you're using them to try to make a difference so that hopefully there's... As you were talking about how in four years, the technology will change. Maybe saying a word like ChatGPT or Claude will feel weird in four years and we'll be like, you can't have that in a book that's published four years from now because it'll look so antiquated. It's kind of sad that some of the systemic issues will probably be the same in four years, if not worse.

Cathy O'Neil: 00:45:53 Absolutely.

Jon Krohn: 00:45:54 So in what ways are you hoping that Orca and Ocean could at least chip away a little bit four years from now and we can be closer to a Star Trek enterprise world? The



- Cathy O'Neil: 00:46:01 Star Trek vision is really very dear to me because it's abundance. It's an abundance mind frame mindset and it's like sharing with other people. Of course, they have infinite resources, so that helps. So Orca, I started when my book came out because I was like, "Oh, people are going to be really worried once they find out that their hiring algorithms could be against the law because of anti-discrimination in hiring laws. And they're going to want to hire me so that they don't get in trouble." And not really. I was imagining and by the way, I'd also, because I'd worked in finance, seen how bad rating systems make things worse. So like the AAA ratings and mortgage backed securities made things worse. They were mathematical lies, but they propped up a loaded housing market. It made it into a huge bubble. We saw the financial explosion of that whole system.
- 00:47:02 And I was like, "We got to do better than that. " So it was like a dual goal of setting better standards for algorithmic auditing and helping people address their risk of getting regulatory pushback or getting sued. I started this in October 2016. I think you can see where this is going. Pretty quickly after that I was like, "Oh, maybe people aren't going to be so worried about that. " And by the way, I'll just back up one second. I was an independent consultant while I was writing weapons trying to make a little money as a data scientist. And I started working for the Attorney General of Illinois on a payday loan case. So it was like consumer fraud. And we forced them to settle. They had this fake insurance product. I'm allowed to talk about this. It's one of the very few things I'm actually allowed to talk about because it was actually settled and it's public, but it was this payday loan company that forced people to pay for fake insurance products.
- 00:48:01 And so everybody who worked there so - called forgot how their business ran once they were deposed on the stand. But because I had all the data, I could sort of reverse engineer the entire business. And that was really exciting.



So I was like, okay, now I can figure out how they treated people and how much this... The usury rate in Illinois, by the way, is like 36%, but the effective APR of these loans was more than 500% after you added this non-optional insurance product. I could also prove that it wasn't really acting as insurance because it was never used, blah, blah, blah. Long story short, there was this killer little graph I built for the judge and the judge was like, "Okay, you got to pay these people back." And then I helped them figure out restitution and that was also very, very exciting. But one of the takeaways there was like, well, if you have the data, you can do a lot.

00:48:58 You can really figure out a lot. And by the way, that wasn't even an algorithm, but it was essentially an algorithm because it was like a company that followed rules. And so it was a bureaucracy. So by the way, when I say auditing algorithms, I'm really auditing bureaucracies. How does this bureaucracy work? How does it treat people? Does it treat these people differently from that? Those kinds of people. That kind of thing. So number one is like, yeah, this can really work. This is a neat thing to do with my skills as a data scientist. That was first. The second thing was like, oh my God, I just went to that trouble of building that awesome graph with that beautiful metric of measurement of harm. It was completely convincing and persuasive to the judge, but now it's gone. It's just completely gone forever. Well, until I get hired by another AG for a similar kind of thing and then I could resurrect this idea, build a new graph with the new data, see if the judge is convinced, blah, blah, blah.

00:49:55 But really what we should do is make this a rule and every lending company should be forced to almost like a public filing, like quarterly statement. Here's the measurement of this metric.



- 00:50:13 That wasn't happening, but by the time I started Orca, I was like, this is the kind of thing that we should be doing in the context of using a hiring algorithm. How do you measure whether the extent to which this is not sexist or racist or ageist or whatever? You can measure that. And why don't you just report that on a quarterly basis? Why aren't regulatory agencies who are in charge of anti-discrimination law enforcement doing that? Well, that's because they don't know how to do that, but I know how to do that. I'm going to show them how to do that. That was the idea of Orca. Didn't really happen. It didn't play out as quickly and it still hasn't played out as quickly as I would hope. But I just want to say that just this week, Jon, I don't know when this is going to air, but just this week, Illinois passed a law that requires auditing of maybe they're AI systems, so maybe not all algorithmic systems.
- 00:51:09 So it's happening. Connecticut recently passed a law as well. And these things are going into effect in 2027 and many of them require auditing. So I'm very excited about that. So that's Orca. And we've been working with lots and lots of law firms, lawsuits, like class action lawsuits, lots of AGs, some federal agencies, but not now because for some reason the federal agencies are no longer working on that stuff with us. But in the meantime, last year after Trump got reelected and our private clients disappeared, because usually we get private clients who are worried about regulatory enforcement are worried about lawsuits. I got the third email from a representative of TikTok asking me to help represent them in a lawsuit that they've been harming young women. And for the third time I was like, "Hell no, I'm not going to represent TikTok because I actually do think they're harming young people.
- 00:52:12 What I really want to do is work for the other side." And then I was like, "Oh, they're never going to approach me because they don't have the money to pay for my services



because I'm not cheap." Although I'm not that expensive either. But it's a lot for an individual who's been harmed to be like, "Hey, Cathy, can you do the data on this? " So that's what Ocean is. Ocean is like a reaction to be like, "Wait a second. We need to be reaching out to advocacy lawyers who are advocating on behalf of somebody who's been harmed and offered to do the data on this. " And so we've been working with like Nitali Jane, who's representing the parents of kids who's killed themselves after talking to chatbot, AI chatbots.

00:52:59 Part of Ocean's work is the podcast, AI Skeptics. I was deposed for a lawsuit against an education company that was stealing student data and scoring them for various things against their parents' knowledge or permission. So that's the kind of work where like just like lawsuits that really need to happen, but nobody's doing the data stuff on it. And when I say doing the data stuff on it, what I mean is this. Somebody's harmed. They know they've been harmed. They're trying to make the case they're being harmed. A lot of the times the pushback from the tech companies are like, "Well, you probably were just getting harmed anyway. We didn't harm you. " How do they prove that they were harmed? And the answer is I come along and I say, "Ask for the following data over time, the trends, blah, blah, blah. What I want to do is get the evidence together that this is happening not just to this kid, but is happening at a systematic level to an entire class of kids who are in a particular situation or whatever.

00:54:07 The point is that when you do, and I know you know this because you're a data scientist, but most of the harm that comes about with respect to algorithms is statistical harm. It's actually really hard to prove a particular person was harmed by a biased hiring system. It could be biased against women, but I could be looked over for a valid reason. So to prove that I'm the one that's being harmed by this biased system is really difficult. But it's



not that statistically difficult if you have the data to prove that it is a biased system. And then to estimate restitution, which is usually to the entire class of people who could have been harmed. Anyway, I hope that made sense and that's what I'm doing with Ocean. And it's really exciting, except I'll tell you what, I've been hustling for 10 years now. I'm a professional hustler for gigs, for algorithmic auditing gigs.

00:55:04 And it turns out it's really easy to offer free work to lawyers who don't have enough resources, but it's a different kind of hustle to get paid for it by charities, like foundations and stuff. So that's what I'm learning now. It's like that's hard.

Jon Krohn: 00:55:24 Regardless of how hard bits of it are, hopefully it'll continue to get easier over time. I don't know. I mean, I hope some combination of societal factors somehow getting better. And I realized I earlier said Star Trek Enterprise World. I actually meant to say Star Trek: The Next Generation, which I mean, Star Trek Enterprise, yes, yes, yes, yes, for sure. But I think you started with the next generation on purpose and I would as well.

Cathy O'Neil: 00:55:52 Most of the ships are called the Enterprise, aren't they?

Jon Krohn: 00:55:54 They are. Yeah. Yeah, exactly.

Cathy O'Neil: 00:55:55 Not all of them. Voyager wasn't. Yeah. But I understood what you meant. It's really about the prime directive.

Jon Krohn: 00:56:03 Exactly. Yeah. And I don't know. I probably naively to some extent. I understand individual issues, but I maintain this, I guess in the '90s growing up, I had the next generation and would watch new episodes as they came out and still love it. Recently have been re-watching all of it to try to have more kind of optimistic ideas. Trying to have the opposite of what you described earlier with Silicon Valley people having consumed all of this -



- Cathy O'Neil: 00:56:37 Dystopian.
- Jon Krohn: 00:56:38 Yeah, dystopian. That's exactly the word I was looking for. Thank you. All this dystopian content. And there's very little utopian content. The next generation is kind of like the only, one of the very few examples.
- Cathy O'Neil: 00:56:53 Jon, I'm so glad you said -----, by the way. It's
- Jon Krohn: 00:56:56 Because I cut.
- Cathy O'Neil: 00:56:57 It felt so lonely over here.
- Jon Krohn: 00:56:59 We still live in this largely kind of dystopian world with algorithms kind of acting against us. And your latest book, which I want to give you just a few minutes to talk about before we wrap up here, your latest book, which came out in 2022 is called The Shame Machine: Who Profits in the New Age of Humiliation? And in this book, you talk about how shame has been weaponized into a destructive, profit driven industry. And it might kind of sound at first glance like that has very little to do with weapons of math destruction or algorithms, but in fact, it doesn't because it's the same kind of idea. Like you were talking earlier and I agree 100% around this idea that algorithm doesn't need to be confined to math or computer science ideas or data science. We can have algorithms effectively encoded into a business process or into a law into a political system.
- 00:57:56 And so in that sense, the same kind of thing. It's like shame is a tool or rather than serving as a tool for community building institutions and algorithms exploit public humiliation and personal insecurities for those classics of financial and political gain that you've been talking about now for over a decade. So I want to give you the chance to talk about the shame machine.
- Cathy O'Neil: 00:58:22 I came up with that notion because I just experienced it personally and so did my children, so did so many other



children. I saw so much resonance with this idea when I talked to young people, especially about their experiences on social media. This is before the chatbots came out. And I'll talk about that in a second because it's related. And things have changed since 2022 when this book came out. But the idea was for me, it's actually not new to make money off of shame. Skin cream has been shaming women for having wrinkles for centuries, makeup.

00:59:05 And it's not just that. It's like weight loss products or are you getting old and senile? You should get this brain medicine. Are you smelly in your particulars? You should buy this deodorant. And that's old school. I would call that old school shame profiteering. But what I noticed was that the social media companies in particular had done something really tricky. Whereas with the old school, they would shame you so that you would buy from them and try to solve a problem that was actually either not a problem or insoluble. So that was like a direct shame, like the direct tactic. We're going to shame you, you're going to buy stuff from us. But social media did something really tricky, which is like, we're not going to shame you directly. We're just going to set up an amazingly perfect platform that is designed exactly so that you'll We'll shame each other and you'll sling mud at each other through shame.

01:00:04 It becomes addictive, it becomes all consuming. And by the way, we'll also exploit you because we'll let those old school shaming products target you with targeted ads. So it was just an amazing, just a framing. I guess really I wanted to do that at the very least to make a framing that people would remember that when we were on social media, we're working for the social media companies when we shame each other. When we're giving into that design, which was made to exploit the way that we skewer each other. And to be clear, backing up a second, shame is not a bad thing. We need shame. And so we would still have apartheid in South Africa if we didn't have shame



because we couldn't appeal to law. The laws were unjust. So the only way you get past an unjust law is through shaming people internationally in that case.

- 01:01:07 Again, because I'm a thought experiment type of girl, I was like, what are the requirements for shame to be valid? When is it a good idea to shame? And when is it a bad idea to shame? And I came up with, for me, something that's been very useful since I came up with it since I wrote the book, which is like voice and choice. Does this person really have a choice about what they're doing? Do you have a choice to have wrinkles? If not, don't shame the person. Also, did they have a voice to defend themselves? It might look on this two second video that they did something terrible. Do you know what actually happened before that? Can they defend themselves? What's the voice? And if you think about it in the context of the constitution, if we accuse somebody of something, they should have defense.
- 01:01:53 They have the opportunity to defend themselves, meet their accuser. None of that seemed to be true for most of the shame trains that we were watching sort of develop. And so that was part of the book. And also just to kind of try to urge people to stop doing the invalid shaming, which is like shaming people without voice or without choice and trying to get more into the valid shaming, which is shaming people in power who should know better and should be making better choices. So people in power have the choice and they have the voice because they're in power. So those are the people that we should be pressuring. And by that I mean our representatives, our policymakers. And in general, I'm a very, very pro protest. It really is exciting to see what happened to the ICE agents in Minnesota and Minneapolis. That was great and that was shame.



- 01:02:49 And thank God for those people. They're heroes. And just one last thing I'll say because I did kind of write that book for young people, like for their mental health.
- 01:03:01 One of the things that I find so disturbing about AI chatbots is that they are in some sense refuges from social media. Whereas social media you're like put on the spot, socially vulnerable, constantly battling and being shamed or being shown other people being shamed. So that's like a marker that you should feel bad. It's like, "Oh my God, I don't want that to happen to me. How do I conform? How do I conform? How do I conform?" That kind of anxiety that it's provoking, it's almost like the big tech companies built that and then they made everyone super neurotic. And then they built these little cocoons of like, "You're safe with me AI chatbots that do almost the opposite, but also in a harmful way." But you know what I mean? They created this terrible place so that they could create this wonderful place. Neither of them is a good place.
- 01:04:05 Neither of them is the right place to develop into a normal person.
- 01:04:12 And I sit on these panels all the time about AI safety and I just get so frustrated to hear the AI folks defend the use of AI chatbots as friends or coaches. I'm like, "Oh, some kids need to practice asking their friends over for a sleepover with somebody who's not going to judge them or whatever." And it's like, "No, actually they need to practice doing difficult things with their friends when they're 13 so that by the time they're 15, they're not as worried about that. " The idea that a chatbot is going to help their anxiety is... Let me put it this way, Jon. When I get my hands on that data, it will not be born out. That's what I'm hoping to do.
- Jon Krohn: 01:05:01 A week ago prior to recording on social media, I announced that you would be a guest on the show. And



of course it was a very popular post and we had lots of questions for you. And so we're going to actually, I've got three questions that I think we can handle pretty quickly and they're transcontinental. So the first one's going to be in North America, then we've got a European one, and then we've got an Oceania one. Amazing. Amazing. And the North American one is actually from Minnesota you were just talking about. And so this is from Sanjeev Weij. I hope I am pronouncing Sanjev's family name correctly. He's a principal architect at a company called Optum in Minneapolis, St. Paul. And he wants to know what we can do about these problems at an individual level or even at a group, community collective level.

01:05:50 What can listeners do to tackle the kinds of problems that you're tackling through Orca and Ocean and that you discuss so much in Weapons of Math Destruction, in the Shame Machine and on your podcast?

Cathy O'Neil: 01:06:03 If I hadn't heard where he worked, I would have been like, "Yeah, that's hard." But I actually went to Optum many years ago when I first started Orca and I was like, "You guys got to be worried about your algorithms being racist in healthcare because of the history of black people not getting good healthcare in this country." And the executives at Optum were like, "Oh, we're not going to worry about that. We're not going to pay you to help us think about that. " And then there was this huge New York Times front page story about racist Optum algorithms, which is one of the case studies that everyone refers to. And I don't know where this guy works, is he a senior data scientist?

Jon Krohn: 01:06:48 Yeah. It says he leads teams in architecture design and delivery of network infrastructure projects.

Cathy O'Neil: 01:06:52 So probably not. But I guess my point is you could start an algorithmic auditing group or lunch meeting at your company and you could say, "What could be going wrong



in the algorithms that we use at this company, especially as it pertains to customers and to patients in the healthcare system?" That's a big question and it hasn't been addressed by this one New York Times headline. In general, what we need is accountability. I think the pushback against data centers is a good thing, but it's not directly related to the kinds of algorithmic harm that I typically think about. I wrote an essay, which hasn't been published yet, but it's like I'm still looking for a placement, but the basic idea is juries of angry parents are our last resort because right now it's, well, maybe Illinois, maybe Connecticut. I'm sure there's constituency pushback against the AI kinds of harms that we've been seeing, especially among young people.

01:08:03 But that stuff is pretty slow moving. It happens at the state level. It takes a year or two to go into force that reinforcement takes another few years. But I do see happening in a faster pace and all over the place, which is very exciting. And one of the reasons I started Ocean is lawsuits. I don't know if you remember a couple of months ago that the lawsuit against social media companies for a particular girl's anxiety, depression, that won and the companies had to pay up. Why did it win when regulation is at a standstill at the federal level? Because lobbyists didn't infiltrate the jury. It's a bunch of parents of kids who they see what's happening to their kids or their kids' friends and say, "Oh yeah, I believe this is happening. I believe this social media company's designed this to hurt young people, to addict them, to keep them on the social media all the time." And I believe that has a deleterious effect.

01:09:02 So I guess what I'm saying is that as an individual in this country, what you can do is go do your jury duty. Talk to other folks in your PTA meetings or other parents or if you don't have kids yet, start getting engaged at a community level about what is this happening to us? We don't actually have to accept this. We can actually change



the rules and we can get people in trouble, like the company's in trouble. We can make them accountable. Because going back to the original discussion we had, Jon, like an algorithm can be terrifying if it's just an unaccountable, unfair system. It doesn't have to be complicated. So how do we insist on accountability? That's through our court system and that's through our elected representatives. So like be involved. This stuff isn't going away. It's all going to happen. It's only growing.

01:10:04 This stuff is only growing. Why do I say that? Because algorithms have replaced every bureaucracy. So all of the things that seem unfair as systems are going to end up being algorithmic questions and algorithmic accountability is a big thing in our future.

Jon Krohn: 01:10:20 It's pretty wild that question asker works at a company that you know so much about.

Cathy O'Neil: 01:10:26 I also wrote about Optum in weapons, by the way, is one of the companies I wrote about.

Jon Krohn: 01:10:32 I hope these question askers don't mind that I'm mentioning their company name

Cathy O'Neil: 01:10:35 On there. Well, they told you.

Jon Krohn: 01:10:37 Exactly. This is all public information. So our second one, it follows on really nicely from everything you were just saying. So this is someone named Mila Vaciuk who is at Deutsche Bank in Berlin. And she's the vice president of group strategic analytics for technology and banking at Deutsche Bank, obviously. And she points out, she says, "We evaluate AI models for bias, fairness and accuracy. Probably not as much as you'd like, Cathy, but there's some of that happening." And she says, "We rarely measure the quality of the human decisions that AI



replaces or augments. Do you think we're holding AI to a higher standard than humans?"

- Cathy O'Neil: 01:11:23 That's a fair question. And I am all for doing both. As I've been saying, I audit algorithms. I audit bureaucracies. And one of the things that puzzles me and bewilders me and not in a good way is how you'll see a company just replace their human bureaucracy with an algorithmic bureaucracy without testing it. Why are you doing that? At the very least, you should have an ongoing AB test to make sure that it is replicating in the ways that you like the things that humans are doing. The example I have from my book from weapons is like the teacher value added model, which was a random number generator pretty much. And so they didn't test it. They didn't compare it to how they used to evaluate teachers. And one of the reasons they didn't do that is because they didn't agree on which evaluation of teachers method was the best.
- 01:12:24 So basically as a way of avoiding a difficult conversation, they're like, "Let's just do something stupid." My opinion is that you should always before you implement an algorithmic solution, you should be running parallel testing of the human version and the computer version. And of course that's going to force you to decide what does it mean for this to be running well whether it's human or algorithmic. But ultimately, the way I do auditing at my firm is we build what I would call cockpits for systems. You wouldn't fly in an airplane without a cockpit. Don't fly a massive algorithmic bureaucracy without a cockpit. Measure everything that you can think of that could go wrong to keep track of it and monitor it and to see if it's going wrong. And that means you have to define what you mean by wrong. So that's the tricky part, but of course you absolutely need to do that.
- Jon Krohn: 01:13:23 Great answer. And I had a feeling you tied into the kind of bureaucracy auditing points that you've been making



earlier in the episode. Final question for you is from Miriam Kakpoor, who is actually a PhD student that I'm co-supervising at the University of Auckland in New Zealand. And so she's a robotics PhD student. And she says that she and her friends who have read weapons of math destruction were discussing this upcoming interview that I have with you. And they had a number of questions, but I'm just going to pick what I think is the one that has been least addressed, which is who audits the auditors? So yeah, if we build a system to check for algorithms for bias, what stops that system from having its own blind spots?

- Cathy O'Neil: 01:14:10 Oh yeah. It's absolutely a thing. I was just on a call earlier this week with folks in Massachusetts who are thinking about writing a law that requires auditing third party auditors. And I was like, "Listen, you got to make the language in that law really strong because there's a bunch of my competitors who will just literally write whatever you want to hear for a fee. And the fee will be low because it's not very hard to do that. And I will charge more because I do a thorough job. So if you're just going based on cost, you're going to go with a cheaper person that's going to give you whatever you want to hear. So we need a better system. I didn't just start Orca because I want to get paid. I started Orca to like improve the actual standards or to actually improve makes it sound like we have standards.
- 01:15:04 We don't have standards. We have to invent standards and we have to make sure they're good. So one of the things I've been calling for is that the methodology has to be open. Even if the answers are private in the context of like legislation saying like you need third party auditors. Every time you have a third party auditor, the third party auditor has to say, "Here's my methodology. I posted it here. Here's my Python code. I'm using this way of measuring the extent to which this is biased or this is working or this is fair." So you have to define your terms.



So that's not the end of the answer because we're not going to agree. M and my competitors aren't going to necessarily agree on what's a good methodology, but at least we're going to have the conversation, which will be an ongoing conversation about like why this methodology?

01:15:54 Why not that? What makes this more robust? And I'll finish by saying, if you really want robustness, you don't just ask one question. You have that whole cockpit panel, a dashboard kind of perspective on a system to make sure it's working. Because you can game a system to make it look good in one way, which is exactly what happened by the way in finance with the VAR value at risk system that was absolutely gamed. And I saw that up close because that's what I worked on in finance for a couple of years. And I was like, oh yeah, if you just rely on one number, it's easy to stuff all the risk in the last top 5%. So you need more than one way of looking at risk.

Jon Krohn: 01:16:45 Brilliant answer, Cathy. Dr. O'Neil, thank you again for an amazing episode and answering all of the questions and all the - Great

Cathy O'Neil: 01:16:52 Questions.

Jon Krohn: 01:16:52 Random tangents that I've gone off on.

Cathy O'Neil: 01:16:54 Thank you for all of your hard work in preparing for this conversation. It's been really fun.

Jon Krohn: 01:17:01 I'm glad you enjoyed it. I'm super lucky to have an amazing researcher named Serge Masis, who's a great data scientist in his own right who makes it super easy for me to prepare for episodes. It's a big shout out to Serge there. Before I let you go, how should people follow you after this episode? Obviously we have the AI Skeptic podcast that they can listen to you on. Is there any place



that they could also find you on social media or anywhere else?

- Cathy O'Neil: 01:17:24 No. I got off of all of it because I don't like working for those people. I still have my blog Mathbab and I sometimes write on it, but mostly I just post my podcast links. And I do really love feedback to my podcast. So that's the best way of probably interacting with it. There's an email address, which I wish I knew on the top of my head, but I will send it to you for feedback on AI skeptics. But yeah, that's the best way of doing it. But thank you. Thanks for having me.
- Jon Krohn: 01:18:03 Perfect. Yeah. We'll have that in the show notes as well as your MathBabe blog. Thank you so much for taking the time. It's been seriously an honor for me, someone I've looked up to for a really long time, for at least a decade that I've been aware of your outstanding work. And yeah, really enjoyed this conversation. Thank you.
- Cathy O'Neil: 01:18:18 Yeah, me too. Thanks.
- Jon Krohn: 01:18:19 Wow, what an episode in it. Dr. Cathy O'Neil detailed why an algorithm doesn't need to be complicated to be terrifying since the real dangers are secrecy, unaccountability and having no way to opt out. How white collar workers whose keystrokes are now tracked are getting a taste of the surveillance, degradation and dehumanization that truckers and teachers have endured from algorithms for decades. How auditing an algorithm really means auditing a bureaucracy, why she wants every algorithmic system to fly with what she calls a cockpit, parallel testing against the human process it replaces, plus a whole dashboard of metrics because any single number can be gamed. And she talked about why algorithmic harm is statistical harm, nearly impossible to prove for one individual, but very provable across a whole class of people. Exactly the data work her nonprofit Ocean now contributes to lawsuits against big tech. As



always, you can get all the show notes, including the transcript for this episode, the video recording, any materials mentioned on the show, the URLs for Cathy's social media profiles as well as my own at superdatascience.com/1013.

- 01:19:25 Thanks of course to everyone on the SuperDataScience podcast team, our podcast manager, Sonja Brajovic, media editor, Mario Pombo, our partnerships team Natalie Ziajski, our researcher, Serg Masís and our founder Kirill Eremenko. Thanks to all of them for producing another exceptional episode for us today for enabling that super team to create this free podcast for you. We are deeply grateful to our sponsors. You can support this show by checking out our sponsor's links, which are in the show notes. And if you yourself are interested in sponsoring an episode, you can get the details on how by making your way to Jonkhrohn.com/podcast. Otherwise, help us out by sharing this episode with folks that would find Dr. O'Neil fascinating. Review this podcast on YouTube or on whatever podcasting platform that you listen to your podcasts on. If you write a review on Apple Podcasts, that is especially helpful for us.
- 01:20:18 So bonus points to you if you do that. Subscribe obviously if you're not already a subscriber, but most importantly, just keep on tuning in. I'm so grateful to have you listening and I hope I can continue to make episodes you love for years and years to come. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science podcast with you very soon.



SuperDataScience