# AI and You

Transcript

Hello, and welcome to episode 92! Today, we'll finish the interview with Ben Shneiderman, Emeritus Distinguished University Professor in the Department of Computer Science, Founding Director of the Human-Computer Interaction Laboratory and a Member of the Institute for Advanced Computer Studies, all at the University of Maryland.  He is a legend in the field of human-computer interface design, and has received *six* honorary doctorates in recognition of that. He is the lead author of Designing the User Interface: Strategies for Effective Human-Computer Interaction, a book I used as my own definitive reference to the topic when I was writing that kind of software. His new book, Human-Centered AI, was published by Oxford University Press in February 2022 and there's a link to it in the show notes and transcript.

Last time we talked about Ben's distinctions of rationalism and empiricism in human-computer interaction, metaphors in HCI, and he drew some fine distinctions for us between some different forms of AI that were built from the human interaction paradigm. This week we'll talk about, among other things, emotions in AI, what it's like to get into this field, and the role of standards in human computer interfaces. Like for instance, just to mention something coming up, Section 508, which is a requirement of the Americans with Disabilities Act that requires government websites have user interfaces that work for people with certain physical challenges, like vision, or hearing. You can hear how precisely Ben parses language when it comes to these issues as we get back into the interview.

So if we draw an analogy with user interfaces, those are much more mature. They've had decades to mature in, and I no longer need to make choices about what color am I going to make this button? Do I need to make sure that this one has the same font as this one? Those are taken care of by libraries that are designed by people who are better at UI than I am. And now in AI though it seems that we're at the beginning of a new field and it's much more Wild West, more like user interfaces were in the early eighties and remembering failures of things like the Therac device where the bad user interface killed someone and that resulted in standards that resulted in things that we could trace through to like 508 compliance, section 508 and other policies, various standards, committees that met. And I wonder what we need, and what do you think we need in that respect to tame AI now that it is in such this nascent state where people are running all over the place creating things but also creating problems.

Very good. I agree. I like your opening phrase of 'maturation' and yes, technologies do need to mature. They go through early stages where it doesn't always work out and we take in the empiricist view of collecting the data to know when things went badly. And so I'm all about that and incident reporting systems would be a good thing. Sean McGregor's work at the Partnership on AI to build the AI Incident Database is a really valuable contribution. There's more than 1000 incidents already reported on that database and that will help us and the insurance companies are

desperate for that information so they know how to set premiums on not only cars, but other kinds of devices. And so we need to have a more open reporting system. So the key thing in response, is transparency and more reporting on data. A second key thing that follows naturally is independent oversight. One of the things that makes civil aviation so safe is the National Transportation Safety Board, that when a plane crashes or train or ship, the right people fly in and begin an investigation, those reports are respected and valued with recommendations that have teeth in them. So, you know, there's a good model. Independent oversight is an important step, and that needs to come to AI. My book and my framework has 15 recommendations of things that need to get done. Five of them are what sort of for engineering teams need to do, like explainable algorithms and testing for bias and fairness and other aspects. And then there's five things that managers need to do to build a safety culture across the dozens of teams they may have that build systems. And then there's five things that go beyond that outside the organization to secure independent oversight. And potentially government regulation for some things. There's a common belief that government regulation will limit innovation, but that hasn't always been the case when done well, such as for automobile safety and automobile fuel efficiency. Government regulations were enormous stimulants to innovative ideas. And so if we have, as the Europeans do have, a requirement for explainability, then we trigger tens of thousands of papers about AI explainability. And that to me is progress. So we'll see how these things play out to get just what you were talking about, the maturation of this technology, so that it's measurably safe, reliable, and trustworthy. And metrics that let us know, agreed on metrics that are transparent, that we have the data that we agree on the way the data is collected and reported and that we have intermediate-sized organizations, Consumer Union, *Consumer Reports*, Underwriter's Laboratory, and the IEEE's current efforts on standard setting, the P-7000 series of standards those are developing now will be another way we get to the maturation you describe. And and my job is to make sure that happens not in 50 years, but 15 years. Okay. That's why I'm out there speaking about it because I want to see these things happen and I want to see that happen soon. And I want these systems to get built and made reliable, safe and trustworthy.

Is it accurate to say my perception is, and I'm asking is this accurate that there's a much larger roll here outside the technology itself in HCAI than there was in user interface design. User interface design resulted in things like consortia that defined libraries or ultimately caused libraries to be built where a lot of the work of that design was encapsulated by people, and then I just got to use it. You have talked a lot about things like governance and ethics and regulation, which are facets that are managed outside of the technology itself and is that an accurate perception? And what does that indicate to you?

You were saying that in HCI, the idea of consortia and standards emerged -- your saying relatively rapidly, you believe?

It seemed to be much closer to the technology. It was industry technologists ultimately designing that technology that would make computers easier to use. And it seems to me that what you're describing in AI involves a lot more people who are not that close to the technology, who may be in government. Tell me if I'm wrong.

Well, I agree that current movement about AI is huge and it's become a worldwide phenomenon. I mean, that's nothing like what I've ever seen. I've just been reading the draft report about the AI index, about the reviews, 30 countries' policies towards AI. It's a 600 page document from the Center for AI and Digital Policy. And it's really impressive what's going on and everything all the way up to the United Nations Declaration of Human rights and as a basis for these things, which I am all in support of, down too much more specific things about how products get and services get developed and reviewed, evaluated, certified and distributed. So there's a much bigger ecosystem, but I would say we may be the point to look at is there's from some sources resistance to this direction. And some of the companies, some of the leading companies, want to maintain their own control over what gets built, rather than engage in these larger consortia. You mentioned before the very the disability rights provisions of section 508 that have done a lot, still a lot to do, but it's pretty wonderful that, you know, your Mac or Windows laptop has these accessibility features and my colleagues at University of Maryland, the Trace Center just celebrated 50 years of doing it. And the W3C at MIT The Web World Wide Web Consortium's Judy Brewer and many others have worked to try to make this technology accessible to a wide range of people and all of that needs to, you know, that kind of strategies need to get put into place. And so we have to move from a world where it's a company just doing it, putting it out there and seeing what happens, towards a more social mechanism that involves so called sociotechnical systems, and that involves insurance companies auditing companies; I'd like to see the large auditing companies that audit financial status also become auditors of AI systems. And so we'll need to put in place these mechanisms that ensure reliable, safe and trustworthy.

To take a tangent here, artificial intelligence has the capability of evaluating measuring human emotion, which is a dimension that I don't think software had before unless it was counting how many times a user pounded on the keyboard in anger. And that seems like an area where the role of HCAI would be particularly interested in the way that we would be using AI that evaluates our emotions. There are so many aspects to that. Is that an area that you spent time in?

Be careful, Peter! Uh yes. Computers have been put to work to try to assess the emotional state of a person. An early theory suggested there were six or seven states joy or anger, fear and so on. And that's turned out to be not adequate. not a reasonable theory of how humans behave. Human emotion is far more complex than that. And the number of times a person pounds on a keyboard is not necessarily a sign of anger. And it means different things for different people under different circumstances. And so there's a shift from, I see the advanced research shifting, from saying", oh, we can determine human emotion" to, "okay, we can measure human behaviors." Okay, so I like Microsoft's recent work. Oh I've forgotten the name of the tool now. But basically it's for meeting management. And instead of trying to guess human emotions, what it does is simply report how often each of the participants was looking at the other people was speaking, how many times they interrupted someone else, how many times they spoke up or didn't speak up, how long they spoke for. And it provides guidance, it provides information for each of the participants so they can adjust their own behavior. I think, you know, trying to gauge human emotion and the galvanic skin response or other approaches, proven to be imperfect approaches here. And so you want to you take into physiological factors muscle skin, you know,

contractions and how people are working, but you want to be careful about what you do with that. And to avoid two things: to avoid this idea of shaping human emotions. Let people shape their own emotions. Or making machines that quote, you know, have emotions. I think that's kind of a cute idea. But I don't see that happening.

And you were talking there about what I would say overselling the idea that we're that AI is over promising along some dimension. And I wonder whether that is something that is inevitable in some circumstances, whether the over selling actually happens without trying. I was looking at a new AI the other day, a new version of GPT-3, that's, now you don't have to just give it examples. You just say what you want like, "write a story about bumblebees and hedgehogs" and it will do it and it knows what you want, all kinds of things and it's of course hard under circumstances like that to avoid thinking that there's something behind the curtain. I wonder what we need to do to change the way that we perceive this to make it safer for us because that's the overarching theme here is what makes it it's safer for us. Never mind some of the other aspects. You've said that it will take a generation to change attitudes and expectations towards a human centered view. Can you talk about how our the way we perceive AI ought to shift?

Yeah, be careful when you say, the GPT-3 will know what you want. GPT-3 is very impressive technologies and you know, it may be useful for some things, but be careful. Be careful what you put it to work for and be careful about what it does. It does very strange things as well that are unexpected. I hope that will improve the time. These people are my colleagues and friends. I want them to succeed. But you know, you have to be careful about what expectations you have. I think as time goes by, we'll have a clearer sense, a diverging sense of what people can do and what machines can do. And I think we'll be getting better and better at designing the machines that do excellent things that empower people, that give them superpowers that enable them to do miraculous things just as the bulldozer or the gives the operator terrific powers that no human has had before or the pilot of the airplane. All those things are pretty amazing ways that amplify the power of individuals. And so that's where I want to get to. I want to get to these powers. I think that the idea of teammates or social robots will slow the progress towards those high performing systems. I don't think photographers think of their camera as their partner or teammate. I don't think that military pilots or tank operators think of their devices as partners. They are in charge. They're operating the device and that's what they want and they're responsible for what happens. They recognize their responsibility. And so I think those are important principles that will become more clear over time and will always have people who promote wild ideas like Flat Earth or you know, or stick to old beliefs, but I think we have a growing consensus of where this movement is. I mean even in the last six months, I'm just extremely pleased to see the growth of interest in human centered AI. Your listeners might, you know, look at Wikipedia on the page on human centered AI, and I was just pleased how many organizations there are, and there's a growing set of activities under that. In your opening, you mentioned the Stanford Human Centered AI group which you know, another impressive collection. But there are about 30 university centers that have human centered in their names, or responsible AI or they talk about some phrase like that and the sprouting of workshops even at

AI conferences -- maybe especially at aI conferences about human centered AI gives me great hope. The NeurIPS conference, the big event, 12,000 or more people attending each year last year had its first workshop on human centered AI that I participated in. They'll do it again this year. The Dagstuhl workshop in Germany this summer will have a human centered AI group. One conference after another special issue of IEEE Pervasive Computing on human centered Ai. Okay, so I've got a long way to go and I appreciate this podcast as an opportunity to let others know of what's happening and I hope more people will join in and come to recognize that human-centered AI is a positive movement which is devoted towards taking breakthrough algorithms of AI and putting them to work in products and services that will really bring benefits to people; that will support human values and dignity and social justice, sustainability. There are all kinds of things we need to think about when we put technologies to work and that's what this movement is about and I'm so pleased that many people are joining. I have run a Google group on human centered AI. put out a note once a week and there are about 2000 people now signed up for that. So that's also a sort of another indicator and I'm pleased that my book Just came out past week. It's leaped to number 33 on Amazon's list of top books on AI. So that's a pretty good start. As I say, there's a lot more to do, but that's where I'm going. I'm here for the duration and I'll be talking about these things.

So if I may ask about then the roles that are available for people who want to be involved more in this, who want to make a career out of it in the way that you had people whose title was UI designer, UX designer, human computer psychologist; what roles do you foresee being developed and where would they happen in businesses, academia, government, education?

Oh, definitely in businesses. That's where we need it and we see it already. I mean, IBM's multiple toolkits of Fairness AI and that's eight or last time I looked tool kits that support explainable AI. And these different features is a clear indication they now call it human centered Ai the google group called People in AI research (PAIR) is another strong indicator of the value and importance. They're a research group but they're influencing the product side as well. there are 50-page guidelines document is the probably the best one to read. It has like we had guidelines documents for user interface, we now have guidelines documents for AI and so that's another positive development. Microsoft Responsible AI Research and their office that supports these activities is another strong indicator. So definitely businesses are the place that this will emerge, that's what we want. I think there's also going to be increasingly in government agencies, we see that happening, the Food and Drug Administration, National Institute of Standards and Technology, one after another are going down the road of – NIST particularly is onto trustworthy AI they've had six or seven workshops and events. I was an organizer of one of them and they are continuing down this path. So all of these things give me hope that we have a brighter future coming.

We've had two people from that IBM center on the show, Michael Hind and Kush Varshney. So to make it concrete: if HR manager of a business comes to you, large enterprise and says Ben, how do I write a job description, a job req for someone that's going to come in and kick start this at XYZ enterprise, what should they be doing?

Great. I mean that's exactly what we're seeing happening already. And there's many different words and terminology, whether they're talking about ethical design, that's a popular older notion. For me, ethics is a good foundation but it's the bridge towards practitioners that are important. We want to build on the ethical foundations but then we want the practical aspects to get done. So you're going to see the language of ethical responsible AI, trustworthy AI, explainable AI, fair unbiased. All those words that describe what they want to achieve will be there and those who understand the methods for doing it, the classical methods of human centered design of user experience design and the newer notions of how to design and build and evaluate them within the context of machine and deep learning. Those are the points to look at. And so I like your approach to identifying what the HR manager needs to put in a job description. And it's those good words that talk about the criteria they want, what they want in the systems and the products and services they will build. Reliable, safe and trustworthy is my Keystone. But there are about 25 terms that have been used.

What are the people in this area at the moment, the people who need to be designing more responsible AI, more human centered AI, what are they either uh most ignorant of or getting wrong the most?

Oh dear. There's a risky one. But I'll take the bait. Facebook. Facebook could do a great deal more to give users control over what happens. Yeah, there are 120 controls embedded deep inside Facebook for privacy and other settings. But it's really tough to make sense of them. But basically you'd want users to be able to control what happens to their data. Iit may not be that individuals do this. It certainly happens when Apple stood up and said, "Hey facebook, you know, we're not going to give you all the data." Recently, the estimate was that cost Facebook $10 billion. And so we will need strong adversaries to deal with powerful companies like, you know, Facebook and Google and to try to control them. I think that's possible. So government intervention is one possibility. And we're seeing that happening with several government actions now. I think the idea of other companies actually taking a strong stance and individuals can have a role. But I think larger organizations, it may take something like a larger group that we join. We sign up for Facebook and we say when I sign up for Facebook, I want to adhere to the rules of the American Civil Liberties Union or whatever group I believe in. I will have them be my editors and if they want to block certain accounts, then I will have them blocked as well. So we may want to appoint intermediaries that are non government organizations or whatever group you believe in. And have them act as your advocate because you can't possibly figure all the things out that are needed to figure to block hate speech to deal with the misinformation, the malicious lies, the terrorists, the oppressive governments and so on. So the movement to take individuals and aggregate them into organizations and have organizations take leadership is what's important, just like Underwriters Laboratory, Consumers Union does for consumers; ther Better Business Bureau. The better AI Bureau. You know, that's what we want.

What would you say to someone who said Ben, I want to do this stuff. I want to be more responsible in the Ai But won't it put me at a competitive disadvantage. Do we need a third party, a standards body to impose that?

Whoa, watch out. I mean, there's a few things going on there, I think for companies, companies that adopt principles that's the 15 recommendations they make that go for a reliable, safe and trustworthy. That will become a competitive advantage. If they can advertise and provide the data that says our systems are safe and safer than our competitors. Hey, it's a competitive advantage. Our customers trust us and you know, that's, that's a very powerful. Ultimately, that happens again, I'm here to make sure it happens faster. So I think that becomes a competitive advantage for people who develop it. For customers, consumers. I think they will want the protections and they will seek the companies that have better business stamps of approval or other kind of agreements that they've adhere to widely accepted principles of design.

You talk several times here about wanting to make this happen faster. That's what you're here for. What are the biggest levers that you can pull that will have the most effect on making this kind of change happen faster?

Oh yeah, there's a lot. I mean there's short term ones and the long-term ones. The short term one is my putting these ideas into a book by doing podcasts and doing talks as I am doing these days to let people to raise awareness is the short term. The longer-term thing is we see the development of training courses for students at universities as well as professionals that will embed these ideas and make sure that they get propagated out because it takes a while for those ideas to happen. The other things are working with government agencies I mentioned, NIST and others that I've worked with and getting them to introduce policies and raise public awareness. Also, journalists are another source of spreading news for the general public and I'm very pleased, John Markoff in the *New York Times* did a very nice profile and my positions now more than a year ago. And so those kinds of placements are what I seek; and your podcast Peter, is my other way of reaching out to get the message.

Well, thank you. Maybe we maybe we will see sometime in the future the Ben Shneiderman Chair of Human centered AI newly endowed at university of your choice.

Absolutely, I hope you'll endow it.

Well, we'll put that out to the listenership for some action here. We will have a link to your book in the show notes and the transcript. Is there anything else you want to draw listeners attention to in your activities?

I think that's a good you know, pointer. The google group on human-centered AI allows them to see what's happening on a weekly basis. The twitter account @humancenteredai just go straight there and sign up to that one. So there's all these ways; visit the Wikipedia page. Tthere's a growing set of ways that people can engage and I look forward to feedback and comments, please get in touch with me.

Fantastic. And they can get in touch with you through?

I'm easy to find. I'm at the University of Maryland and ben@cs.umd.edu gets me.

Fantastic. Ben Shneiderman, thank you so much for coming on *AI and You*.

Thank you, Peter Scott.

That's the end of the interview.

In today's news ripped from the headlines about AI, Nikkei Asia reports that Japan's National Police Agency is set to create a permitting system for the use of level 4 self-driving cars for transportation services in rural areas. The government aims to put level 4 automated driving systems to practical use in areas, aimed mainly at elderly passengers, by the end of the fiscal year ending in March 2023, expanding them to more than 40 locations nationwide by around 2025. Operators will be required to have remote monitors and staff on call to respond immediately to incidents. As we've talked about before, when it comes to the deployment of level 4 or 5 autonomous vehicles, everything depends on where it happens. There are some environments where it's possible, but the huge majority of locations are not going to work for that kind of automation.

Next week, my guests will be Hannah Grubbs and Shea Sullivan from the Institute for Digital Humanity, a bi-partisan, cross-cultural, and, importantly, student-run digital ethics think tank out of Minneapolis, Minnesota.

That's next week on *AI and You*. Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

**http://aiandyou.net**