

FINISHED FILE

EDUCATIONAL DEVELOPMENT CENTER
SARAH MACGILLIVRAY
OCTOBER 3, 2019

CYBERLEARNING 2019 WEBCAST
9:45 AM CT

Services provided by:
Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
800-825-5234
www.captionfirst.com

This text, document, or file is based on live transcription. Communication Access Realtime Translation (CART), captioning, and/or live transcription are provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This text, document, or file is not to be distributed or used in any way that may violate copyright law.

>> As we switch computers, also we're not going to take questions as they said before. There is time kind of around lunch and stuff to kind of use that for question asking if that sounds good to you. Our next keynote is by Dr. Safiya Umoja Noble. She's an associate professor at UCLA in the department of African-American studies. You might know her as the author on racist and algorithmic bias that was entitled algorithms of impression how search engines enforce racism. She has a PhD in labor rare researchers. I'm a little bit partial to her. Her academic research focuses on the design of digital media platforms on the Internet and their impact on society. Her work is both sociological and interdisciplinary marking the way digital media impacts issues race, gender, culture and technologies. As you know, our Cyberlearning research is interdisciplinary. It brings together the learning sequences and security science. He want to make more essential the humanities and other social sciences that explore how technologies impact humans in different ways. Her talk today if we can get it up on the machine is entitled the problems and perils of harnessing big data for equity and justice. And we're restarting. That's fun. It's all good.

>> SAFIYA NOBLE: I have a couple jokes that I've workshopped. It's not ready yet. It's not you. It's me. I saw that a couple times in my life.

>> Thank you.

>> SAFIYA NOBLE: Thank you very much for that lovely introduction. As I was walking up, my computer crashed like it does. So I will take a minute to tell you a little bit about myself and maybe I can just get like a grace period of a few minutes at the end. I am really excited to be here. I was over at the Smithsonian this morning and there are a lot of people there gathered also talking about similar sets of concerns, mostly having to do with things like digitization of knowledge. I come from the field of library and information science. So in that field it's always interesting to be among scholars listening to people talking about knowledge and education and information and the way we talk about it in our field in LAS is quite different because we're feeling so many dimensions of the threats of misinformation and disinformation that are really all around us. So some of the things I was so interested in the previous speaker, Mike, in your talk because these concerns about data and the making of data and the access to data and information appears as though if all data and all information are equal. Or stable or reliable.

Of course, if I get a talk together here, I'm going to maybe generate some provocations around that (chuckling) Apple says, sorry, can't read that. Can't read that PowerPoint. So we'll see what happens here.

I'm -- I wrote a book called the algorithms of oppression. In it I was really interested in the ways in which I had spent 15 years in advertising and marketing before going to grad school. And as I was going to back to graduate school, mostly to atone from myosins from having booze and cars and stuff like that in advertising. That's

what that was about. It was interesting to me it was a moment where when we were the height of the Google book digitization project, which some of you might remember. I know some of you are also like, sorry, we did that because that was happening.

Everyone was talking about Google like it was the new public library. And that was interesting to me because as I was leaving advertising, we were hiring so many freelance programmers to come in and help us game the system and gain search to bolster the visibility of our clients' products and services. No one thought for one second in advertising that Google search was the public library. It was an advertising platform as far as we were concerned. I think also as far as Google was concerned. So it wasn't even a mystery.

So it was interesting to kind of enter that juxtaposition where everyone was engaging with Google as if we were -- that's my sign I'm ready. Engaging as if the new kind of commercial digital technologies were going to be some type of -- did you see how I did that? I amaze myself that I kept that going and got that up.

(Applause)

>> SAFIYA NOBLE: It's funny to me. Please don't tweet my mess out.

(Laughter)

>> SAFIYA NOBLE: It was amazing to me that people were so, just enamored with search in particular. So a lot of my work is focused on search, because I find that people use it as kind of their first pass or first go-to when they're trying to learn about something. Of course, this is incredibly dangerous. I'm going to talk to you about why that is dangerous in the short time we have today. Then also think about -- I want to also think about -- I'm going to show you some failures of what happens when we're hyperreliant upon, you know, big data driven platforms that are of a commercial nature in particular. Then also think about how so many of these kinds of impulses are getting replicated into educational technology and learning technology spaces and see if we might imagine some other possibilities.

Now, it was funny when I was writing -- funny by terrible. When I was writing my dissertation and I was looking for four people who wanted to be on a dissertation committee because I wanted to argue that algorithms are bias at the core level of code but also that technologies were not tools, that they were more than tools, that they were also a set of values, and a that they were political. It was very difficult in 2010, '11, and '12 to find four people to put their names to such a dissertation. Here we are now, everyone is like the bias algorithms. I'm like, are they? Because remember when I was saying that a minute ago and everyone was saying that's impossible because it's math and math can't be racist. We've come a long way baby.

(Laughter)

>> SAFIYA NOBLE: I'm very happy about that because once we start to create new normals, we get to have different conversations. And I think that is a very exciting moment that we're in, that we are

in a position to finally really main stream critique of various types of digital technologies and that means that we can do something different, better alternate.

You might remember for like the three people in here who are familiar with my work. The first study that I did was looking for different types of identities in search. I was -- this is important just as to get us all on the same page about the implications of big database technologies. Starting in about 2009, all the way through 2012, I was collecting searches on all kinds of identities, black girls, African-American girls, women. I used the Census categories and I paired them through gendered categories, boy, girl, man, woman, and so forth. I did an analysis to see how are people represented in these spaces.

One of the things that was so disturbing was to see that black girls in particular as well as La Tina girls and Asian girls were hypersexualized and exclusively represented by pornography. How can this be? People would say things to me like, well, you know, this just reflects what the public wants, what's out there. Search engines, it's not Google's fault. Search is just a reflects of the public and how the public thinks. I was like, no, that's not it. That's an insufficient explanation because if that's the case, then any community that is in the majority will get to control the narrative about everybody else forever.

So we can't really rest on that this is just a matter of that. Also, let me tell you that girls in this context is women. Because the subjects and the agents that are a part of the porn industry are grown women. Where do the girls, where do the children, where do the vulnerable get to be represented? They get to be -- this is like in my gender studies class this would be sexism 101. Women get coded as girls. Go, fix, solve. These are some of the assumptions that are made about the characterizations and classification of information about people.

And I collected a lot of information and data and evidence about how these kinds of processes happen over and over and over again. That's in a book. You know, for years I was talking about this, and people would say, one of the things I also try to parse is the commercial, kind of political economy of search engines of industries that have the most money also win. Vulnerable communities, adolescents are never going to compete with the amount of money that the porn industry has.

Of course, a lot has changed since 2012. Certainly large commercial platforms are doing a lot more to try to manage the types of content that move through their work and also the criticisms have engendered and kind of forced a lot of different kinds of review. People would feel sad when I would give this talk years ago about this specific research. I'm in the same kinds of logics of misrepresenting people and information and ideas, let's say through a presidential election. Now people are really interested in talking about the way in which algorithms work and inform the public. Here's another quick search. This was a tweet in fact that went viral about

when you get a search on something like professional hairstyles for work. I can imagine any one of my undergraduates getting ready to go on the job market, looking for professional hairstyles for work.

When you did unprofessional hairstyles for work, you were given almost exclusively black women with natural hair. I like to where my unprofessional hairstyle to work and is the hairstyle I wear every day. When you look for professional hairstyles for work are exclusively white. White sisters it's the pony tail that makes it professional. And the white helps.

(Laughter)

>> SAFIYA NOBLE: These are the things that I think people were not paying attention to, quite frankly, or they were just kind of in spotty ways noticing. So I tried to amass all of the evidence that I could find into this book.

Now here's a more sophisticated and nuanced phenomena that happens on line. When white nationalist take over a specific type of disciplinary logic and want to reframe it. Here you have anthropology and for the two anthropologists I've ever met in my life who understood the politics of what this meant, it's really important because there are a number of ways in which ideas about people and communities get hijacked and put in service again of things like Holocaust denying, anti-Semitism, and blatant kinds of white supremacy. Not that unprofessional hairstyles for work isn't also a manifestation of racist ideologies but you have coordinated efforts to destabilize the knowledge basis. I think this is important that we not take for granted the sources of knowledge that one might be drawing upon as they're developing all kinds of different learning models and sources and resources that they're using.

So here we are. We're at a moment in time where we are experiencing what the Financial Times in 2019 declared the word of the year, techlash. Part of this is from the Cambridge Analytica, data mining, broke, and selling data to persuade us in various political inclinations and, of course, for other industries, people working in advertising, the whole paradigm is based on buying and selling and trading our data. It was like, Cambridge Analytica is just one of thousands of companies that are harvesting information about us, trading and selling it and making data profiles about us. I was in a conference a couple years ago with a director of the FBI, don't ask why, it was weird to me too, and how we ended up on the same spot. And he said, as far as the state was concerned, as far as the government was concerned who we are, who you are is your data profile.

And I wonder how many of us even know what our data profile is. Who are we in the context of large-scale systems that are somewhat interoperable and in other ways are not that are collecting and amassing a sense of who it thinks you are. I will tell you that after years and years of doing searches on various types of genders and a speaker and writer about racist and sexist bias in search, I didn't stop getting the porn sites when I was looking for black girls until they were finally removed. This level of customization and

personalization also that we think is happening to really accurately hone in on who we are is also not true. And we've seen some really important studies out of Europe in particular where we see that the ideas about personalization are really about aggregating us, like if I were at a library conference, who is the dog people with me. It would be like three people and everybody would be the cat people. It's like trying to figure out who are the cat people and who are the dog people and selling us to platforms. So this is one of the other dimensions I think that we have to think about when we use machine learning and this kind of data that gets scraped about us or that we buy about students or that we're collecting on students and on various publics to somehow create predictive analytics about them.

I think that, you know, we are just beginning to see the importance of the conversations about the harm that come from predictive analytics. Of course most of these data profiling systems are in fact about that type of work. Now these are some people that I think you should know about. You like how I put myself up there but I didn't put myself first.

(Laughter)

>> SAFIYA NOBLE: That somehow makes it better to me. These are people who have been writing about the kind of consequences of many forms of technologies. People have been writing counternarratives, if you will, about the kind of in response to the tech know Utopianism in response to the digital divide. I think about Craig Watkins writing about how African-American teenagers were also the most adapt at using digital technologies and were really like engaging in very sophisticated ways with tech at a time when people were labeling kids of color are digitally divided and technologically deficient. So the deficit modeling that happens about people of color has really been documented and history sized and studied differently.

These are some names I want to put on your radar in terms of dipping in and thinking about a whole host of different types of technologies that you might be engaging with. And, again, their broader social and economic consequences for vulnerable communities.

Now a few perils because why not? We're on a role. This is a talk best served right before the coke tall hour.

(Laughter)

>> SAFIYA NOBLE: So I'm hopeful you'll get a bite and move past some of these difficult conversations. So my colleague Jonathan Ferner at UCLA wrote an important piece about a decade ago about classification. He is a philosopher. He rights about library and knowledge classifications. He's really an expert in knowledge organization. I think that when we talk about things like learning and learning science, it really has to be in relationship to knowledge and knowledge organization and the making of knowledge. So if you don't know a lot about that, I invite you to come to our conferences in the field of information studies in library and information science, because I think there would be some good ideas to double click on there.

One of the things that he says that we need, you know, to kind

of respond to these systems -- again, one of the things that people forget about search engines is that when (?) and Larry page were working on search, they were borrowing from the LIS practices of citation analytics. Now, you know, if you're in computer science, you might think like citation analytics, these are great because scholars credit people that are important. Every time they cite out that's a signal that it's legitimate. The whole system of hyperlinking was really based on the logics of citation analytics.

See, if you're in the field of library and information science, you know that citation analytics, we have a deep body of work that problematizes that process. For example, there are some people who might get cited as their work being bogus or terrible. But they got cited. Awkward.

(Laughter)

>> SAFIYA NOBLE: There also might be people who have become so -- their concepts and the ideas they've introduced have become so mainstreamed that they don't get cited anymore because it's just common knowledge. So if we talk about power in class, nobody's citing Marx behind that, for example. So these are the kinds of things that I think are also, when we think about the design logics in architectural logics we use, we have to reach out into other domains and start thinking more critically about what's at stake. Of course, one of the things we know is there's a tremendous amount of bias in classification systems. I would love to some day get to talk to you about that. But I don't have enough time to go too deeply.

One of the other things we know is there's this adherence to the notion of neutrality. The classification systems or the data sets that we use are neutral. We teach quite a bit in our department -- we don't just teach data science but the social construction of data. The data becomes representative of very specific rich experiences but they get in some ways abstracted from the minutia, the details of what they're representing. At some point they start to get aggregated in such a way that they start to not make sense. Yet, we use these models -- we use this data to fuel the building of our models.

I think one of the things that I think is just something we can't forget is that data is always a representation of something that happened in the past. It's something we've collect the. It's something we've documented. It's something we have been able to assess. And that's means if we use data that's made from the past and we use it in our predictive models about mow students will behave about which of the courses of new technology that's coming online now, we're just predicting the past into the future. We want to ask ourselves at what cost will that come? How much of the past do we want to see predicted into the future?

Matthew Reid wrote a great piece about algorithmic bias in library discovery systems. I think it's a caution air tail about the mechanics for those of you who build systems discovery systems and other types of interfaces into large databases like a radio library database, let's say, he really carefully traces how that happens. I would definitely recommend you take a look at his work. Tara

Robertson also wrote an important piece about the digitization of a really important lesbian porn magazine that ran from 1984 to 2004 called on our backs which was a magazine that was designed for a very closed, intimate community that knows each other pretty well. And then their collection was digitized and made available to anybody with a web connection, a browser, who could find it. What's at stake when we digitize things that are not meant for the public too. How do we discern. All data is not available or maybe shouldn't be subject to mass availability. These are the kinds of things where many people have been put their careers and their livelihoods put in jeopardy.

In the book I document, for example, for like a teacher, a 70-year-old teacher who someone found her racy photos from the '50s and she lost her jobs. She's like I'm not racy anymore. I'm 70. It's cool.

(Laughter)

>> SAFIYA NOBLE: These are the kinds of things when we think about, what does it mean to just collect everything? BlandChet writes about the social value of forgetting. There's a reason why we forget and it's why juvenile records, for example, are sealed and don't follow you into your future so you can have a chance at a future. What happens when we collect, we use these new models that we're developing, especially in educational context -- I shutter to think that my 18-year-old self and the nonsense papers, quite frankly who's writing a good paper when they're 18 would follow me into my career? The private relationship of feedback and engagement that I'm trying to have with the faculty what taught me over the course of my career or have taught you over the course of your career, the experimentation with ideas that you can only do in that type of exchange between you and your instructor, the person you're learning from, that being digitized and put into a system to a system like turn it in and commercialized and sold at some point and then being data mineable. These are the kinds of real projects that are happening right now.

I think there's a lot at stake that we are -- we have not properly thought through. And I'm attempting to a partner with you in some of those conversations. Ultimately, you know, we like to make apps and we like to think some of the problems and concerns that we have are going to be solved. I appreciate the previous speaker's comment about the tech. I could have wore my shirt that says tech won't save us. Join a union. Sometimes I think I'll get in trouble and I might get jumped at an airport especially like San Jose. So I keep it low.

I will tell you that thinking about the kinds of tech overlays that we're doing on various publics, I think we are -- it's so -- we have yet to see what will happen when students in ten years, students who documented their whole lives and for whom we have documented them and participated in building systems that document and model them, what happens when they want to run for senate? Or Congress? They want to be a social worker. We are putting a lot at stake when we don't think about the ethics and the issues of justice that we're going to face through the hyperdataification at various -- many publics are not in a position to even respond to or opt out of these

systems. I break down how the Moodle works for all of my students at the beginning of the quarter. I say, so I'm going to be able to see when you tell me yes, professor, noble, I did the readings. I can see that you haven't actually logged in for 16 days. I want to make sure that all of your other instructors can do that. They're like --

(Laughter)

>> SAFIYA NOBLE: The oh, my God, the watery eyes, you know, response. These are the kinds of things that are not transparent even to the students that many of us are working with and care about so deeply. Ultimately I want us to think about in places where my work is going. I'm really thinking about herbs of power, intersectional power. Obviously I care about things like races and gender and class and power and how they work in digital technology systems. It's very important. But the other things we need to think about as we rush toward the hyperdataification of everything, it has a cost.

There was a great study that just came out that's buzzing around Twitter right now from the MIT tech review that looks at a study about how large-scale data modeling programs can emit, as you can say, as much carbon as five cars in their lifetime. There's a lot of conversation about this on Twitter. Think about the unbridled dataification that's happening. It's not without impact. Not just on the environment. I spent quite a bit of time last summer looking at e-waste. Where do devices go with the constant upgrades? What do we think we're getting? One of the things I can tell you is we're getting huge toxic e-waste neighborhoods communities around the world where the digital waste goes. The life -- the estimated life span of people who work in those toxic waste sites taking apart all of your devices and trying to recycle and reuse is in the 30s.

So these things happen in places that are not in the United States for the most part. I think we have a lot at stake. If we look at the extraction, the mineral extraction and mining industries, of course, there's incredible exploitation happening there.

So I think we have a very -- we are the consumers pushing for these kinds of adoptions. I think if we're going to do that, if you're going to do that, then more knowledge needs to be put into the equation and thinking about how do we offset, how do we mitigate, how do we resolve, how do we restore these types of investments that are being made and how do we recover what is lost and who is paying the price? How do we address who is paying the price for these things?

I wrote about these with my colleague Sarah Roberts. We wrote about what's at stake for students specifically in the context of educational technologies. I think this might be a piece that could be of interest to many of you. Again, thinking about the information communication technology infrastructures that we are building. Of course, I know we're gathered here to talk about Cyberlearning. And one of the challenges that I think we often faces that thinking about ethics, thinking about issues of justice are often additive, add on at the last second, throw a word in there to say that we're doing something ethical. I would implore us to rethink whether we want

issues and concerns about ethics and justice and society to be additive or afterthoughts. I actually participated in a recent NFS grant. You know, I was invited to participate because of my expertise. It was the grant -- it was more like \$300,000. It's a good small proposal. The ethics part was maxed out at 50K. I was like, that is interesting in terms of like our orientation to these very real sets of concerns. I think whether the call asks for us to address these kinds of concerns that are very specific to the context in the places where you all do your work, we don't have to wait for the calls to think about these things and make them more central. We can just design our work with a type of care.

So I want to leave you with a few things we can think about. I think we can consider the effect of hyperdatafication in the context of things I shared so briefly. I'm sorry there isn't more time to do more.

Resist making these ideas an afterthought or additive. I think we must protect vulnerable people from surveillance and data profiling. I've been interviewed by a chronicle of higher he had. Where a campus is using a geolocate type of tech where they're basically surveilling students all over the campus. How much time are they in the library. How much time are they here, there, everywhere on the campus? It's sold as a tool of helping students learn better.

So you guys already know what I said.

(Laughter)

>> SAFIYA NOBLE: And you could just ask students what they want. You don't have to put a GPS tracker on them. You could ask people in the library, how are people using the library? You could engage with all kinds of people on the campus, faculty, resident, advisors, the people in the dorms and so forth, how people are doing. We don't have to track them and then -- I see these as just many sites of profit making out of our institutions, unfortunately. But I think that there's a lot at stake, especially for students when we don't think through more critically what we're doing. We also need to fund digital media research. If there are any program officers here and you want to talk about that, talk with me. I'll be here for a minute. There are a lot of people in this room that care about this. But we don't have big bodies of public funding available to look at these kinds of issues. Quite frankly, it's very difficult to do this type of research, those of us who do it, and we need more resources.

I also think we need to be curating the indexable web. I say this to librarian professionals. We need more democratic supported pathways to knowledge. The front door can't just be a commercial search engine. And libraries, academic libraries in particular are only available to people who have a university ID and access.

So what we see is that -- I tell this to policy makers when I'm in DC talking about my work. It's insufficient to regulate big technology companies. Of course, we need that. It's insufficient just to do that and simultaneously defund every important research organization, public research organization that's funding research,

defunding public libraries, defunding higher education, defunding public education.

So the democratic counterweights in terms of knowledge in fact the MIT tech review article I shared a few minutes ago, one of the things researchers say most of the big data lives in private commercial operations. It's becoming increasingly more difficult for academics to do the research to do because it's harder and harder for us to get access.

Ultimately, you know, I think we should never give up. These things are daunting in some ways but they're things we can work on together. And I really believe that our commitment to people and to the environment can eclipse the commitments to the tech. So I'll leave it there and say thank you for letting me speak with you today.

(Applause)

>> Well, thank you, Mike and Safiya for excellent, thoughtful keynotes. I wish we had time to take questions. We do have lunch and that's a great time for questions and discussion. So I'm in the position of having to stop you from going to lunch for just a few more minutes.

I'm sure your head is full of tensions and ideas to talk about. I want to give you a quick heads-up about the afternoon session. I'd like the folks who are leading expertise exchanges this afternoon to kind of make your way up here. We would like to ask you to give quick summary of your session. You can read your abstract or act it out or improvisational dance, whatever you want to do.

(Laughter)

>> And then Quinn will introduce the working sessions for this afternoon. So have we got anybody with -- okay. Great.

>> I'll do the dance.

>> Come up here. Just quick 30 seconds or a minute.

>> All right. So we're going to do an expertise exchange on --
(End of session.)

(Webcast session concluded at 10:55 AM CT)

This text, document, or file is based on live transcription. Communication Access Realtime Translation (CART), captioning, and/or live transcription are provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This text, document, or file is not to be distributed or used in any way that may violate copyright law.
