Algorithmic Justice

ENCS 393: Social and Ethical Dimensions of ICTs

Day 10 – June 8, 2020

Announcements

- Reflection Essay #3 due this Friday, June 12th
- Last readings and final mini-assignment on Monday, June 15th
- Last Zoom class on Wednesday, June 17th: brief conclusion and project questions/workshopping
- Projects due Friday, June 19th

Final Project: Critical Technology Assessment/Re-Design

- Choose an ICT as the basis for your project: this might be one specific technology (e.g. the iPhone 11) or a category of technologies (e.g. smartphones)

- Through research, identify a way to improve this technology that would make it more ethical, more equitable, or more just.

- Create a 5-minute video or presentation that describes:
  - Which aspect of the technology you would change
  - Why and how you would change it
  - What factors aside from the technological design itself are important to realizing this change (e.g. laws, regulations, advertising, intended users etc.)

- Refer to both course material and external sources to justify your arguments

- Work alone or as a team of two

Bias

- Bias is a prejudice against or in favour of a thing, person, group, or idea.

Bias

- Bias is a prejudice against or in favour of a thing, person, group, or idea.
- Biases can be conscious or unconscious (i.e. sometimes we recognize that we are biased, and sometimes we don’t).

Bias

- Bias is a prejudice against or in favour of a thing, person, group, or idea.

- Biases can be conscious or unconscious (i.e. sometimes we recognize that we are biased, and sometimes we don’t).

- Many different types of bias exist! Some examples are
  - **Confirmation bias**: Ignoring or filtering out opinions that contradict our existing worldviews
  - **In-group bias**: Exhibiting favouritism towards members of a group that you belong to
  - **Selection bias**: Drawing (generalized) conclusions based on a non-representative dataset

Bias, Algorithms, and Information Processing

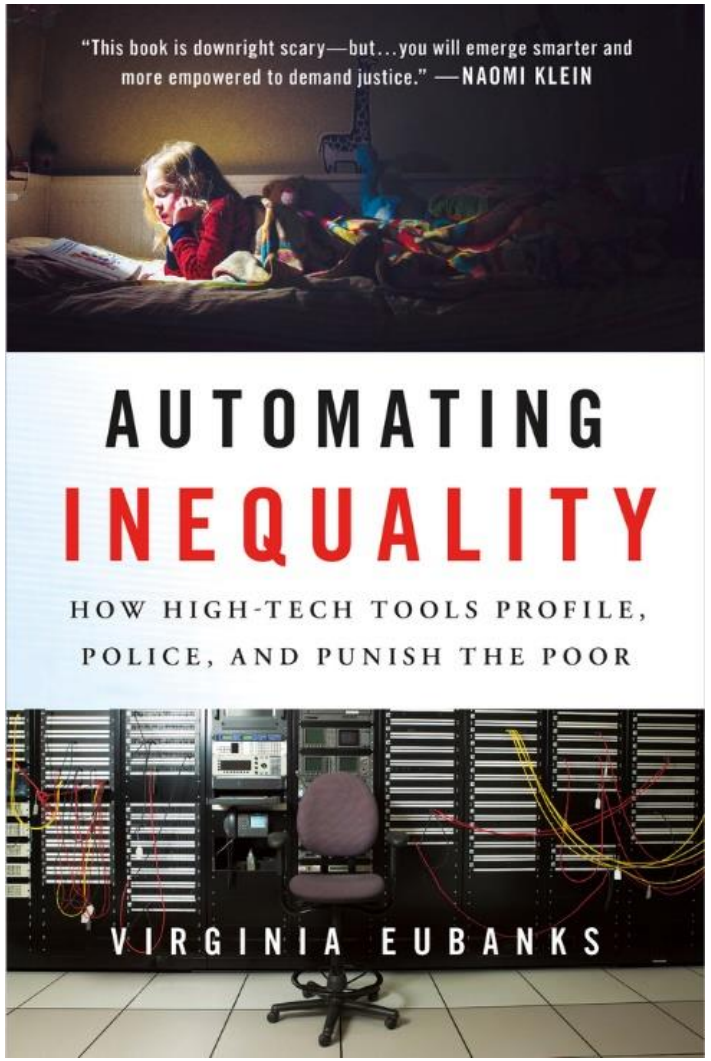
When bias becomes encoded in technologies (even unconsciously or implicitly), it can introduce inequality into a system or reinforce existing inequalities.

Today's readings showed many different examples of biased ICTs.

- Some biases are unconscious
- Some biases are conscious, but have unintended effects

Researchers and advocates for **algorithmic justice** aim to understand and limit the negative effects of bias in technological systems.

Virginia Eubanks, *Automating Inequality*



Virginia Eubanks, *Automating Inequality*

Alleged and True Causes of Poverty.

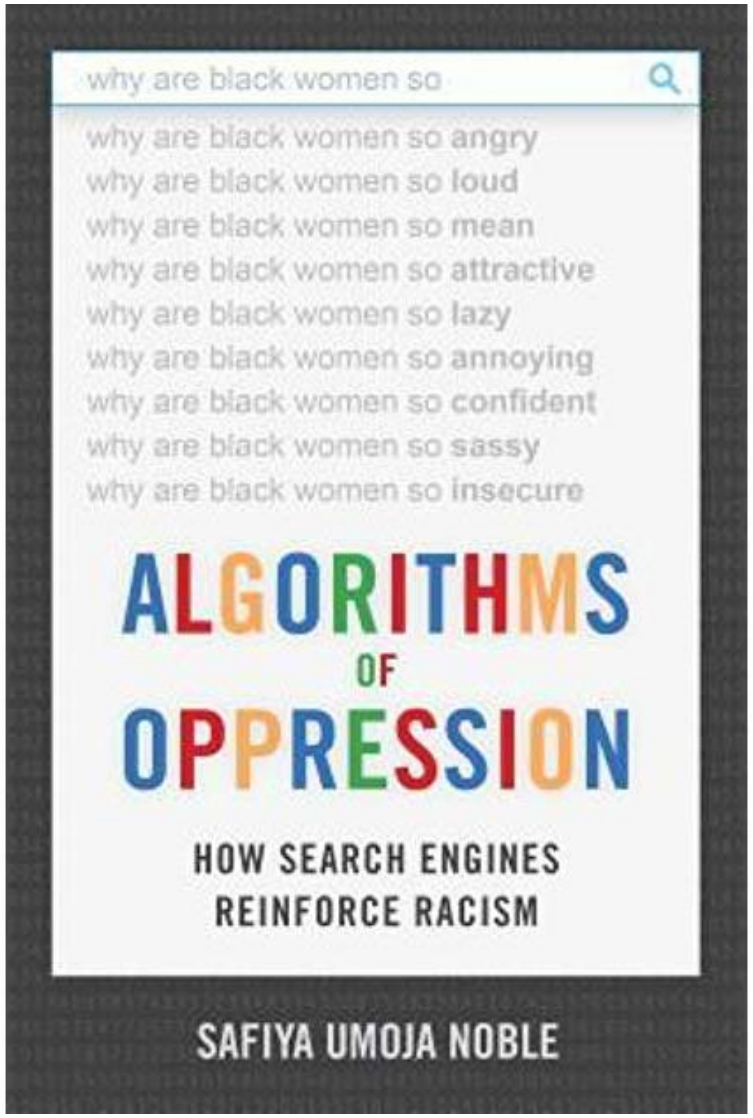
800 cases, C. O. S., New York, 1896–1897.*

	Lack of Employment.	Sickness.	Intemperance.	Shiftlessness.	No Real Need.	Various Others.
Cause alleged by Applicant . .	313	222	25			240
Cause as determined later by Charity Agents	184	164	166	101	121	64

* Lindsay, N. C. C., 1899, p. 872.

Virginia Eubanks, *Automating Inequality*

Safiya Noble, *Algorithms of Oppression*



Safiya Noble, *Algorithms of Oppression*

Noble argues for the ethical importance of Google's search engine:

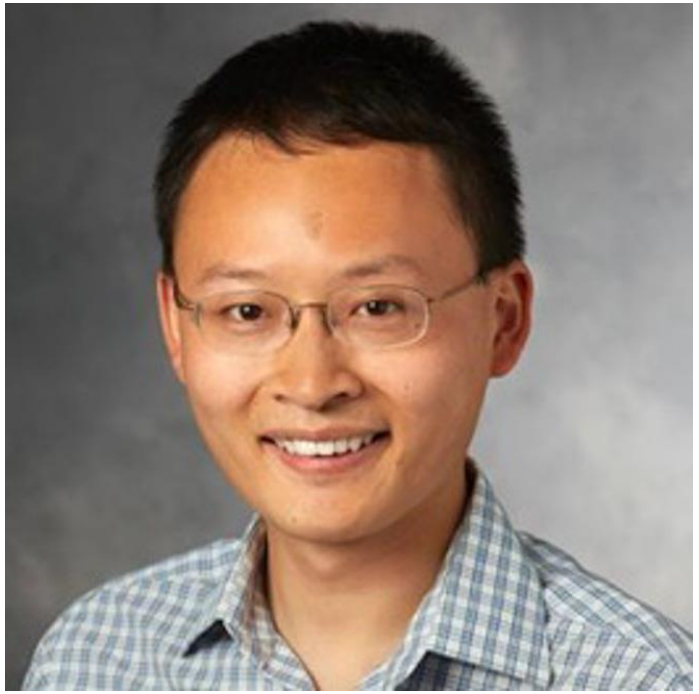
- Ubiquity (i.e. it's everywhere): recall Moor's argument about "logical malleability"
- Trust (i.e. we believe it): recall "invisible complex calculations"
- Misunderstanding (i.e. most people don't really know how it works): recall "invisible programming values"

Safiya Noble, *Algorithms of Oppression*

Noble wants to disrupt some dominant myths about search engines, and Google search in particular, including the ideas that…

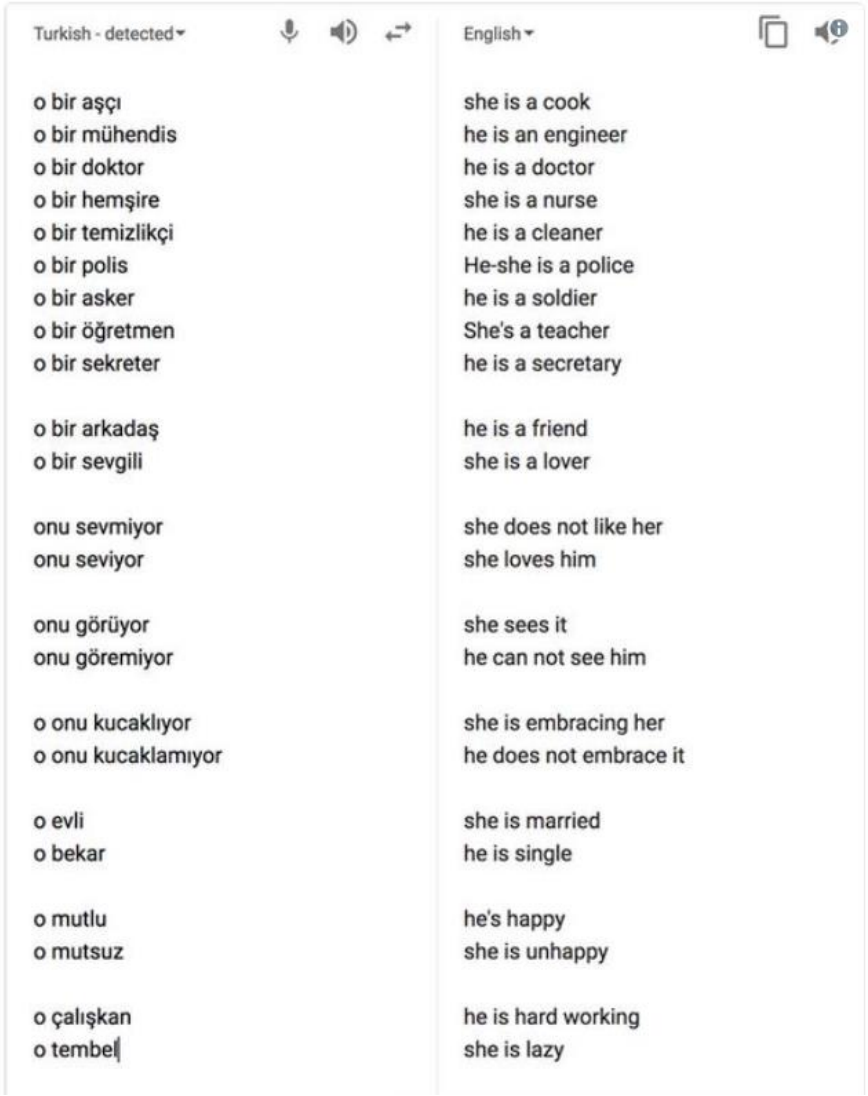
- Search results are objective
- Search results display what is most popular
- Categorizing and displaying information in this way in unavoidable, i.e. no alternatives exist

Zou and Schiebinger, “Design AI So That It’s Fair”



Skewed Datasets and Selection Bias

Reinforcing or Amplifying Bias



Turkish - detected	English
o bir aşçı	she is a cook
o bir mühendis	he is an engineer
o bir doktor	he is a doctor
o bir hemşire	she is a nurse
o bir temizlikçi	he is a cleaner
o bir polis	He-she is a police
o bir asker	he is a soldier
o bir öğretmen	She's a teacher
o bir sekreter	he is a secretary
o bir arkadaş	he is a friend
o bir sevgili	she is a lover
onu sevmiyor	she does not like her
onu seviyor	she loves him
onu görüyor	she sees it
onu göremiyor	he can not see him
o onu kucaklıyor	she is embracing her
o onu kucaklamıyor	he does not embrace it
o evli	she is married
o bekar	he is single
o mutlu	he's happy
o mutsuz	she is unhappy
o çalışkan	he is hard working
o tembel	she is lazy

Reinforcing or Amplifying Bias

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Towards Fair AI

Zou and Schiebinger suggest several ways to improve the fairness of AI and machine learning systems:

- Make training datasets better
  - Make sure they are representative
  - Be wary of convenient binary classifications (e.g. “black/white”) that fail to capture complexity of identities

Towards Fair AI

Zou and Schiebinger suggest several ways to improve the fairness of AI and machine learning systems:

- Make training datasets better

- Label and annotate datasets
  - Label with standardized metadata
  - Include precise information about how the data were collected and annotated
  - Organizations/academic journals can help by changing their requirements

Towards Fair AI

Zou and Schiebinger suggest several ways to improve the fairness of AI and machine learning systems:

- Make training datasets better
- Label and annotate datasets
- Make algorithms better
  - Reduce dependence on attributes like income, race, gender
  - “Nudge” the model to make sure its performance is equitable

Towards Fair AI

Zou and Schiebinger suggest several ways to improve the fairness of AI and machine learning systems:

- Make training datasets better
- Label and annotate datasets
- Make algorithms better
- Use machine learning itself
  - Conduct an “AI audit” to identify, quantify, and eliminate biases

Towards Fair AI

Zou and Schiebinger suggest several ways to improve the fairness of AI and machine learning systems:

- Make training datasets better
- Label and annotate datasets
- Make algorithms better
- Use machine learning itself
- Collaborate
  - Work with social scientists, humanities scholars, and others with relevant social expertise

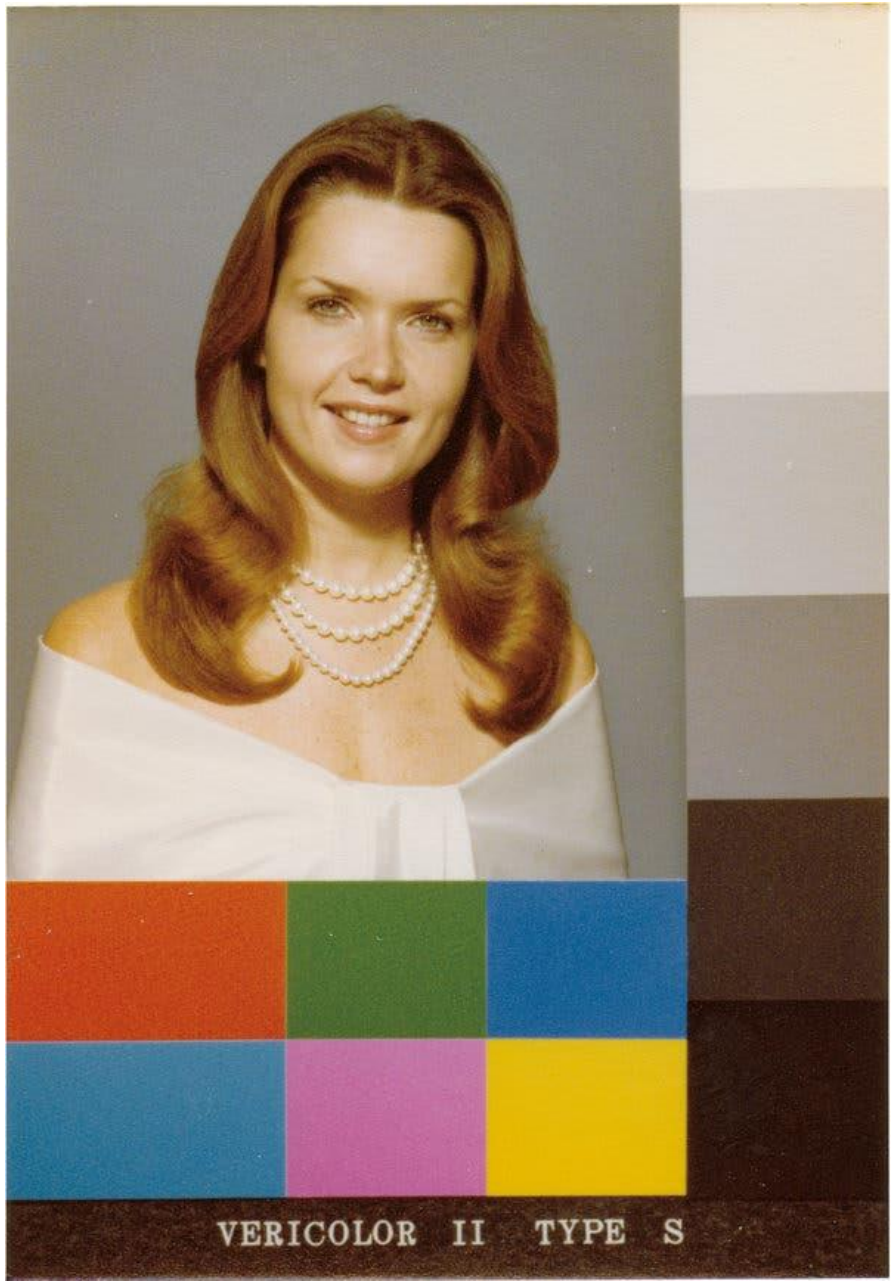
Ruha Benjamin's *Race After Technology*

The "New Jim Code" is Benjamin's term for the ways in which discriminatory designs can encode inequity by amplifying racial hierarchies; ignoring and thereby replicating social divisions; or by aiming to fix racial bias but ultimately doing the opposite.

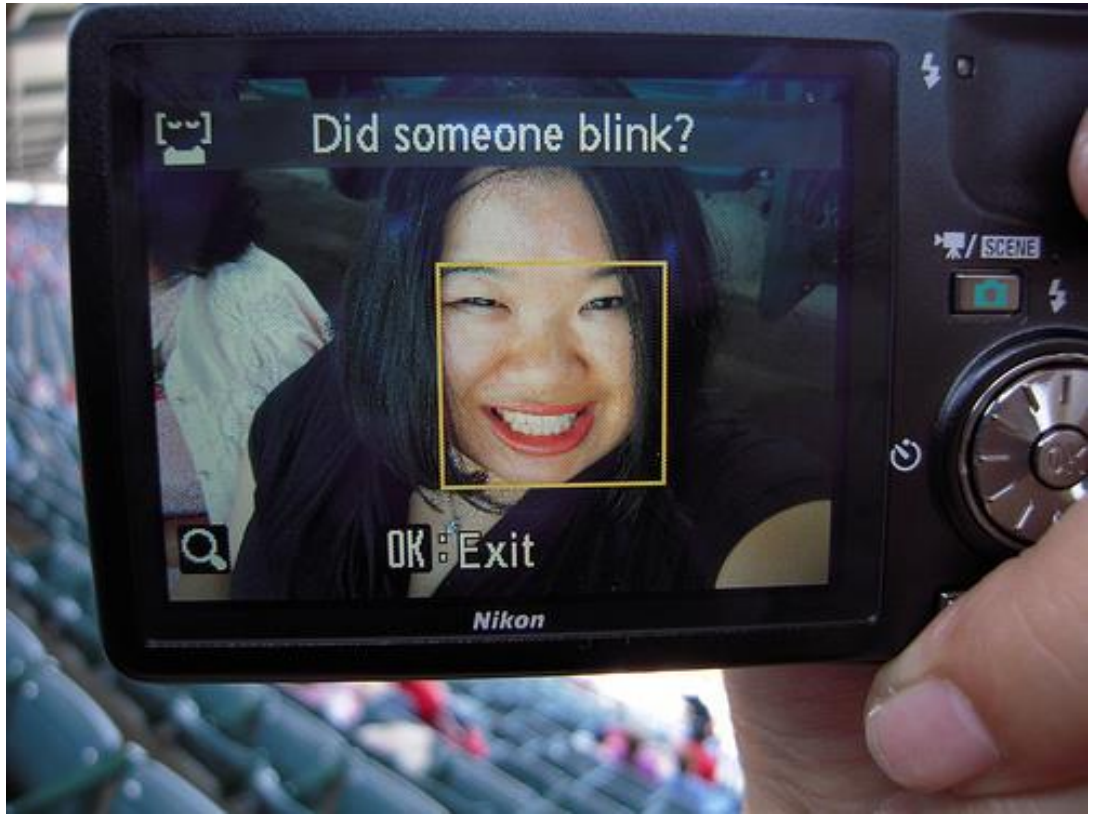
Benjamin's *Race After Technology*

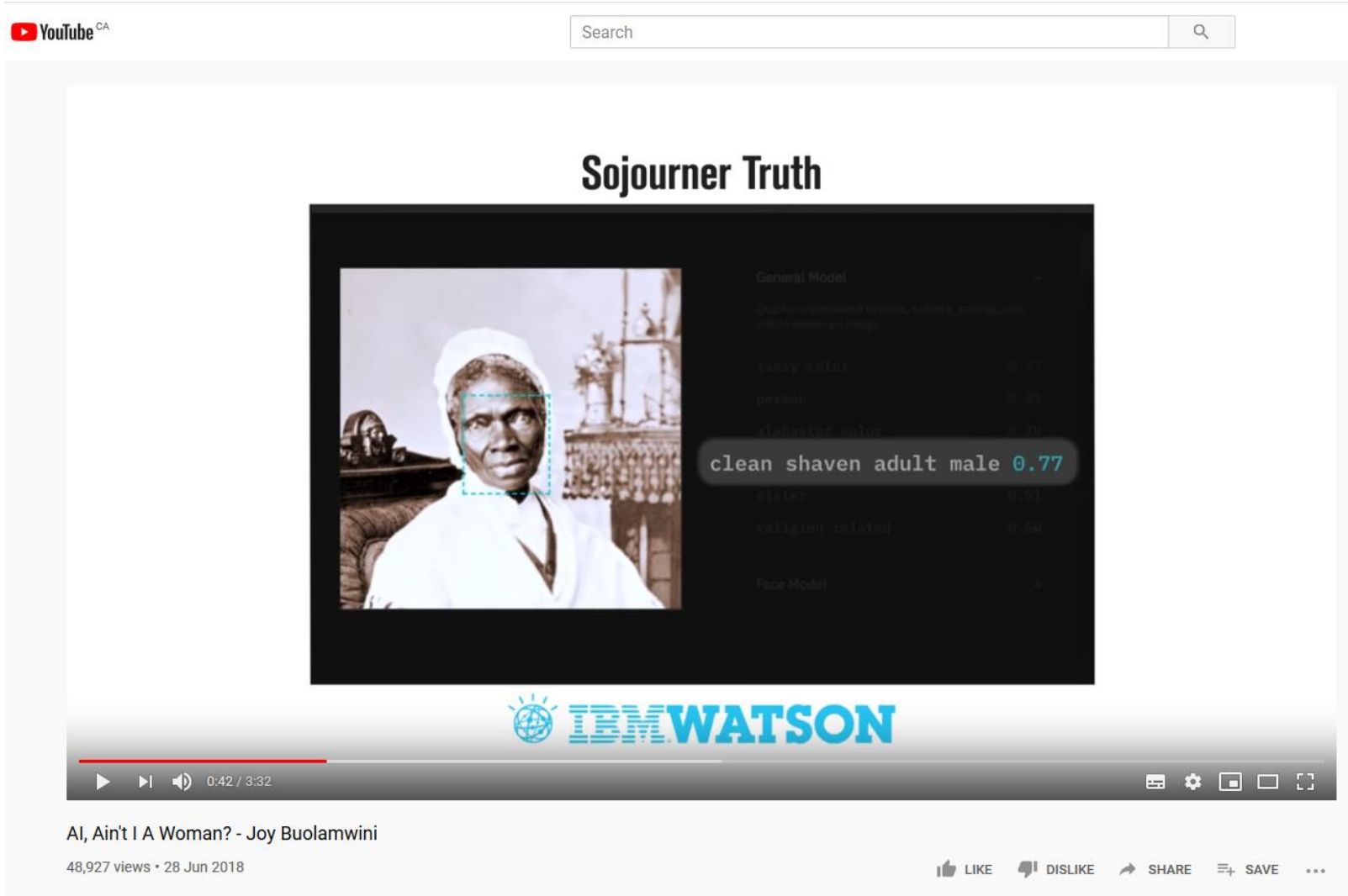
In the assigned chapter, Benjamin explores multiple meanings of the term "exposure," and in doing this highlights different ways in which technologies encode racial biases.

- She first shows how some technologies have been developed or configured to render certain races invisible, or "underexposed."



- “Shirley cards” were produced by Kodak from the 1950s to the 1990s, and were a vital part of film exposure methods.
- Until the 1990s, the cards featured only white models. The film exposure process was standardized around these cards and images, leading to the development of technological processes that did not accurately portray darker skin tones.
- Benjamin demonstrates how such technologies rendered people with dark skin “invisible,” both literally and figuratively


Did someone blink?
OK Exit
Nikon
SCENE



Joy Buolamwini, “AI, Ain’t I a Woman?” https://www.youtube.com/watch?v=QxuyfWoVV98

Benjamin's *Race After Technology*

In the assigned chapter, Benjamin explores multiple meanings of the term "exposure," and in doing this highlights different ways in which technologies encode racial biases.

- She first shows how some technologies have been developed or configured to render certain races invisible, or "underexposed."

- She then demonstrates how inclusion, or increased "exposure," is not always beneficial. (Is it ethical to build a more functional, more "inclusive" facial recognition software if that software is used to target racialized minorities? To develop a precise method of DNA testing if test results will determine refugee status?)

Algorithmic Justice League

Towards Algorithmic Justice

The Algorithmic Justice League and its affiliated researchers aim to “shift the AI ecosystem towards equitable and accountable AI.”

This represents a shift from emphasizing *moral algorithms* – trying to define specific rules and processes that will result in ethical outcomes – to focusing on the *ecosystem* in which those algorithms are created.

Justice and Technological Design

In general, justice-centered approaches to technological design explicitly aim to:

- Increase human rights
- Increase opportunities and resources for marginalized groups
- Reduce risks and harms for all involved

Justice-centered design approaches require:

- Community engagement and nuanced, “contextual listening”
- Recognizing the structural conditions that give rise to community issues and needs

(adapted from Leydens, Lucena, and Nieusma, “What is Design for Social Justice?”)

AJL Core Principles

- Affirmative Consent
  - Real choice for everyone in how and whether they interact with AI systems
- Meaningful Transparency
  - Understanding of what AI can and cannot do, and how AI systems are created and deployed
- Continuous Oversight and Accountability
  - Robust policy that continually monitors AI systems, limits abuses, and holds companies and other institutions accountable for any harms
- Actionable Critique
  - Constructive, research-based criticism, advocacy, and education aimed at shifting industry practices

Mini-Assignment #8: Ethics and Justice

This Mini-Assignment asks you to compare justice-centered approaches to computing with the ethical computing standards developed by the ACM.

Review the ACM Code of Ethics and Professional Conduct that we discussed earlier in the course. Compare the ACM's ethical principles to the core principles of the Algorithmic Justice League, and *identify one substantive way in which these two sets of principles are different*.

Describe the difference that you identify and post your response in the Moodle forum.

This assignment is due before our regular class time on Wednesday (2:45pm).

“Together, we will work to decode the powerful assumptions and values embedded in the material and digital architecture of our world. And we will be stubborn in our pursuit of a more just and equitable approach to tech – ignoring the voice in our head that says, “No way!” “Impossible!” “Not realistic!” But as activist and educator Mariame Kaba contends, “hope is a discipline.” Reality is something that we create together, except that so few people have a genuine say in the world in which they are forced to live. Amid so much suffering and injustice, we cannot resign ourselves to this reality we have inherited. It is time to reimagine what is possible. So let’s get to work.”

Ruha Benjamin, *Race After Technology*

Reading Hints for Wednesday's Class

ICTs, Postcolonialism, and Globalization

- *Rohan Deb Roy, "Decolonise Science: Time to End Another Imperial Era"*
  - A historical summary, written for a general audience and published online.
  - It is more focused on science than on technology or ICTs, but it provides a good overview of some of the themes that we'll discuss.

- *Martin Nakata, "Indigenous Knowledge and the Cultural Interface: Underlying Issues at the Intersection of Knowledge and Information Systems"*
  - You may find this article a bit dense: there's a lot of set-up that is necessary before Nakata gets to the arguments that are most relevant for this course.
  - Focus on understanding what he is saying about information systems in particular.

Reflection Essay #3 due on Friday!