CS 585: Computer Vision and Ethics

Photo Manipulation

Taking a page out of Stalin's playbook, The cover of the Rolling Stones album *Rarities 1971-2003* has Bill Wyman removed from the original photograph, because Wyman left the band in 1992.

Frédo Durand, Ethics and Computational Photography, 2019
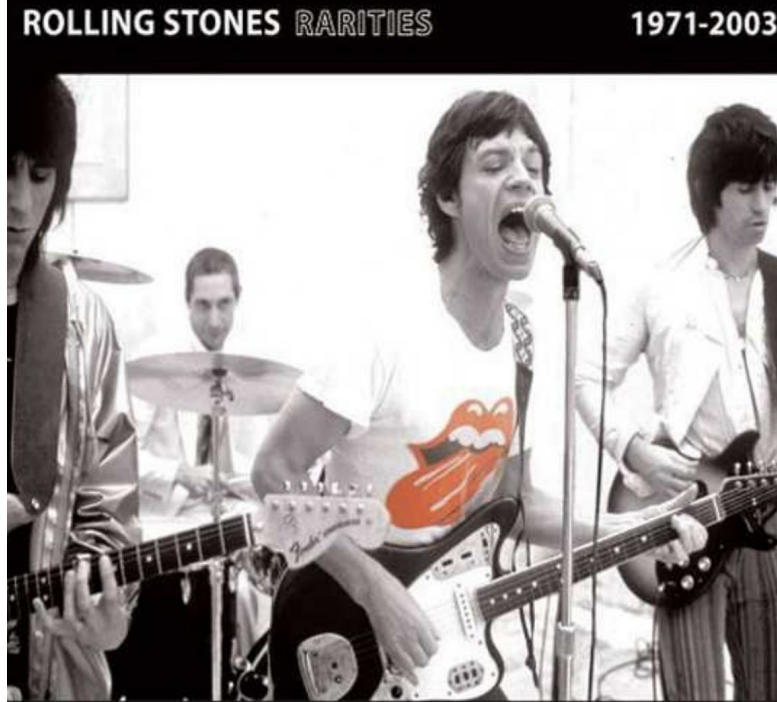
In this 1944 photograph of Benito Mussolini...

...the horse handler was removed to make it look more heroic.

CRACKED.com

Frédo Durand, Ethics and Computational Photography, 2019

Stalin had torturer Yezhov removed

Exhibit: Commissar vanishes

Frédo Durand, Ethics and Computational Photography, 2019

Racist fake fake

Jack Cashill, author of the book *Deconstructing Obama* claimed to have discovered a photoshopped photo of President Obama and his grandparents sitting on a New York City bench...

Presenting this version of the "original picture" below as proof that the photo had been manipulated.

Well, one of them has... but it's the supposed "original". Anyone even slightly familiar with Photoshop will see the stone background to the right of this picture is one of the worst examples of "clone stutter" ever produced. Also, he forgot to Photoshop out Obama's knee, between the grandparents.

CRACKED.COM

Real picture but presumably fake

Fake picture

AP Code of Ethics

- https://www.csus.edu/indiv/g/goffs/135%20photojournalism/associated%20press%20ethics%20code.pdf

- AP pictures must always tell the truth. We do not alter or digitally manipulate the content of a photograph in any way.

- The content of a photograph must not be altered in Photoshop or by any other means. No element should be digitally added to or subtracted from any photograph. The faces or identities of individuals must not be obscured by Photoshop or any other editing tool. Only retouching or the use of the cloning tool to eliminate dust on camera sensors and scratches on scanned negatives or scanned prints are acceptable.

- Minor adjustments in Photoshop are acceptable. These include cropping, dodging and burning, conversion into grayscale, and normal toning and color adjustments that should be limited to those minimally necessary for clear and accurate reproduction (analogous to the burning and dodging previously used in darkroom processing of images) and that restore the authentic nature of the photograph. Changes in density, contrast, color and saturation levels that substantially alter the original scene are not acceptable. Backgrounds should not be digitally blurred or eliminated by burning down or by aggressive toning. The removal of "red eye" from photographs is not permissible.

- When an employee has questions about the use of such methods or the AP's requirements and limitations on photo editing, he or she should contact a senior photo editor prior to the transmission of any image.

- On those occasions when we transmit images that have been provided and altered by a source — the faces obscured, for example — the caption must clearly explain it. Transmitting such images must be approved by a senior photo editor.

- Except as described herein, we do not stage, pose or re-enact events. When we shoot video, environmental portraits, or photograph subjects in a studio care should be taken to avoid, misleading viewers to believe that the moment was spontaneously captured in the course of gathering the news. In the cases of portraits, fashion or home design illustrations, any intervention should be revealed in the caption and special instructions box so it can¹t be mistaken as an attempt to deceive.

Beauty Filters on Social Media

- Smoothing skin texture; minimizing fine lines and blemishes

- Erasing under-eye bags

- Erasing naso-labial lines ("laugh lines")

- Application of virtual makeup, such as lipstick or eyeshadow

- Slimming the face; erasing double chins

- Enlarging the eyes

- Whitening teeth

- Narrowing the nose

- Increasing fullness of the lips

- Highlight Eurocentric features: lighter eyes, smaller nose, and flushed cheeks

- Contribute to social media users' feelings of body image insecurity ("filter dysmorphia"), even plastic surgery

Source: Wikipedia

BOSTON UNIVERSITY

Bias in AI Image Generation: MIT Graduate Asked AI Image Generating App "Playground AI" to Make Her Headshot More Professional -- It "Whitewashed" Her Instead

give the girl from the original photo a profes linkedin profile photo
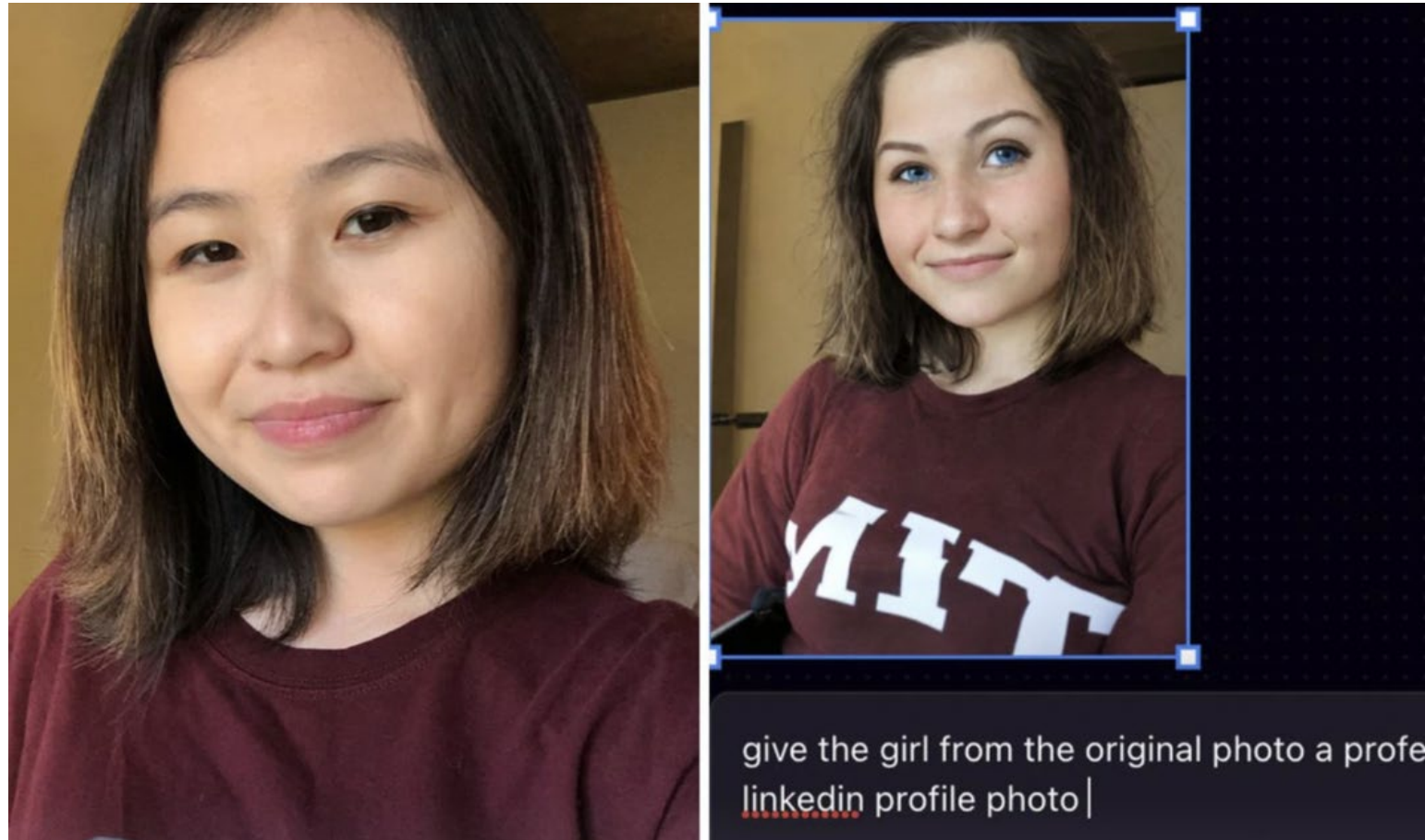
August 9, 2023

Credit: Peopleofcolorintech.com

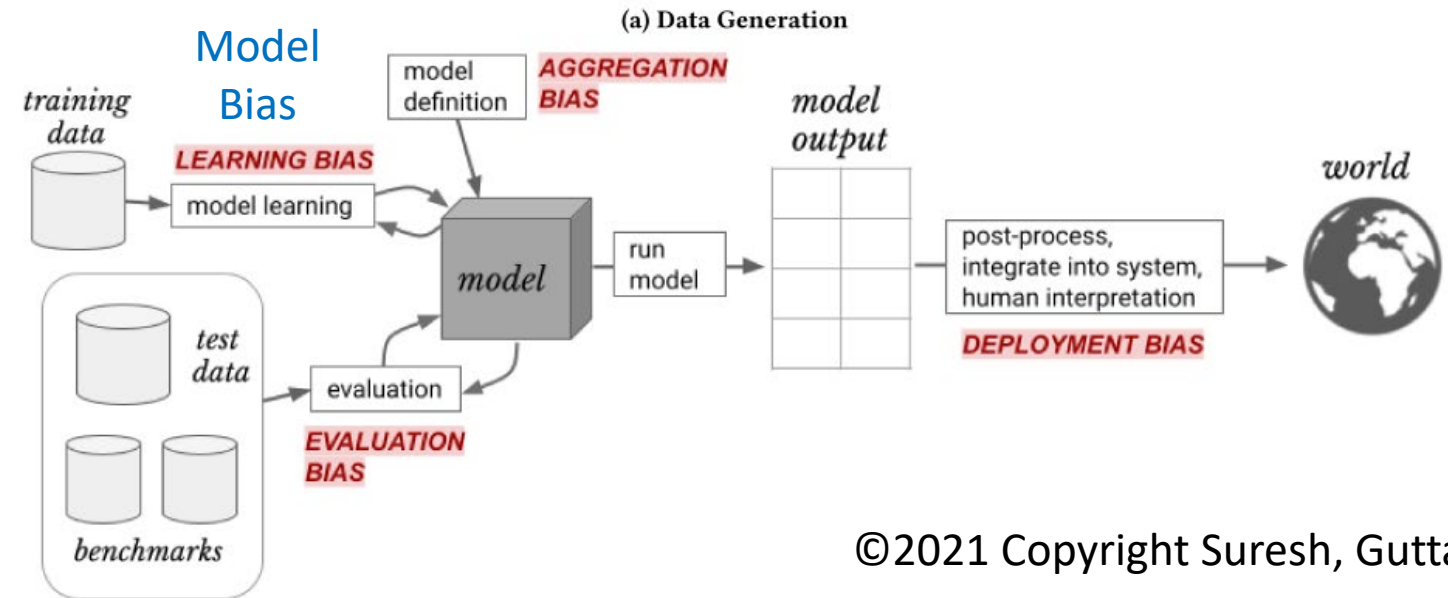**CS 585: Image and Video Computing, Margrit Betke, 2024**

A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle
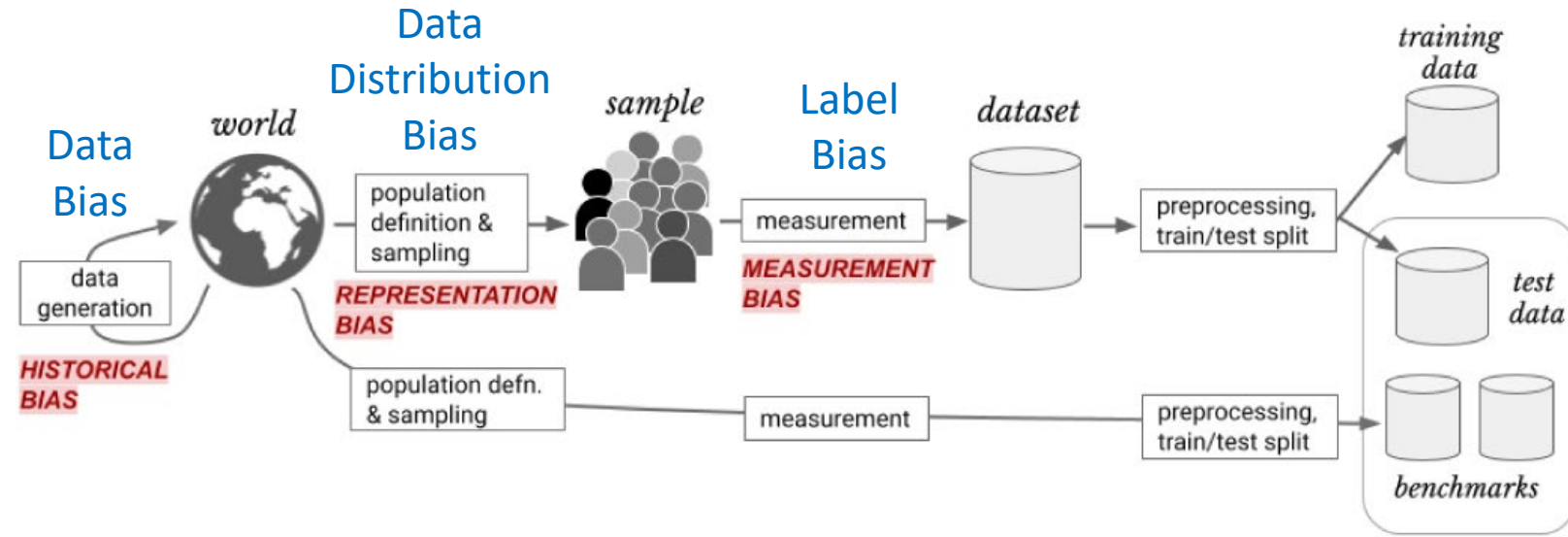
[The 7 Sources of Harm in ML](#)
Suresh and Guttag

ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, 2021

Data Bias · Data Distribution Bias · Label Bias

world → population definition & sampling → sample → measurement → dataset → preprocessing, train/test split → training data / test data / benchmarks

HISTORICAL BIAS · REPRESENTATION BIAS · MEASUREMENT BIAS

(a) Data Generation

Model Bias · AGGREGATION BIAS

training data → model learning → model → run model → model output → post-process, integrate into system, human interpretation → world

LEARNING BIAS · EVALUATION BIAS · DEPLOYMENT BIAS

(b) Model Building and Implementation

©2021 Copyright Suresh, Guttag

CS 585: Image and Video Computing, Margrit Betke, 2024

Historical, Representation, and Measurement Bias in ImageNet

- K. Yang, K. Qinami, L. Fei-Fei, J. Deng, O. Russakovsky, [Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy](#), Conference on Fairness, Accountability, and Transparency, 2020

Problem 1: Annotation vocabulary, WordNet'98 and its insufficient filtering of "offensive", "derogatory," "pejorative," or "slur" labels

Problem 2: Overreach in attempts to label non-visual concepts, e.g., can a "philanthropist" be visually represented?

Problem 3: Lack of image diversity, leading to race and gender bias

BOSTON
UNIVERSITY

Problems 1 & 2:
1: Annotation vocabulary, WordNet'98 and its insufficient filtering of "offensive", "derogatory," "pejorative," or "slur" labels

2: Overreach in attempts to label non-visual concepts

Unsafe (offensive)	Unsafe (sensitive)	Safe non-imageable	**Safe imageable**
n10095420: \<sexual slur\>	n09702134: Anglo-Saxon	n10002257: demographer	n10499631: Queen of England
n10114550: \<profanity\>	n10693334: taxi dancer	n10061882: epidemiologist	n09842047: basketball player
n10262343: \<sexual slur\>	n10384392: orphan	n10431122: piano maker	n10147935: bridegroom
n10758337: \<gendered slur\>	n09890192: camp follower	n10098862: folk dancer	n09846755: beekeeper
n10507380: \<criminative\>	n10580030: separatist	n10335931: mover	n10153594: gymnast
n10744078: \<criminative\>	n09980805: crossover voter	n10449664: policyholder	n10539015: ropewalker
n10113869: \<obscene\>	n09848110: theist	n10146104: great-niece	n10530150: rider
n10344121: \<pejorative\>	n09683924: Zen Buddhist	n10747119: vegetarian	n10732010: trumpeter

BOSTON UNIVERSITY

Solutions for Problems 1 & 2:
    1: Annotation vocabulary, WordNet'98 and its insufficient filtering of
        "offensive", "derogatory," "pejorative," or "slur" labels
    2: Overreach in attempts to label non-visual concepts

Unsafe (offensive)	Unsafe (sensitive)	Safe non-imageable	Safe imageable
n10095420: <sexual slur>	n09702134: Anglo-Saxon	n10002257: demographer	n10499631: Queen of England
n10114550: <profanity>	n10693334: taxi dancer	n10061882: epidemiologist	n09842047: basketball player
n10262343: <sexual slur>	n10384392: orphan	n10431122: piano maker	n10147935: bridegroom
n10758337: <gendered slur>	n09890192: camp follower	n10098862: folk dancer	n09846755: beekeeper
n10507380: <criminative>	n10580030: separatist	n10335931: mover	n10153594: gymnast
n10744078: <criminative>	n09980805: crossover voter	n10449664: policyholder	n10539015: ropewalker
n10113869: <obscene>	n09848110: theist	n10146104: great-niece	n10530150: rider
n1034412?: <pejorative>	n09683924: Zen Buddhist	n10747119: vegetarian	n10732010: trumpeter

Remove

Don't use for
model training

BOSTON
UNIVERSITY

Solution to Lack of image diversity, leading to race and gender bias?

- Adding images
- Removing image

Yang et al., 2019

Solution to Lack of image diversity, leading to race and gender bias

- ~~Adding images~~   Which? Search engines are biased. Diversity was already attempted by ImageNet creators.

- Removing image:

Web interface that automatically re-balances the image distribution according to a user-provided target distribution

ImageNet Examples of "Programmer" and Yang et al.'s balancing results

Original:

Balancing gender:

Balancing skin color:

Balancing age:

Bias in Image Captioning Datasets

E. van Miltenburg, [Stereotyping and Bias in the Flickr30K Dataset](#), Proceedings of the Workshop on Multimodal Corpora: Computer vision and language, 2016

Image comes with 5 crowdworker labels:

1. A blond girl and a bald man with his arms crossed are standing inside looking at each other.

2. A worker is being scolded by her boss in a stern lecture.

3. A manager talks to an employee about job performance.

4. A hot, blond girl getting criticized by her boss.

5. Sonic employees talking about work.

Bias Reduction in Image Captioning Models

L. Hendricks, K. Burns, K. Saenko, T. Darrell, A. Rohrbach, Women Also Snowboard: Overcoming Bias in Captioning Models, ECCV 2018

Equalizer Model

looks at person instead of contextual cues to make gender-specific predictions

Baseline: *A **man** sitting at a desk with a laptop computer.*
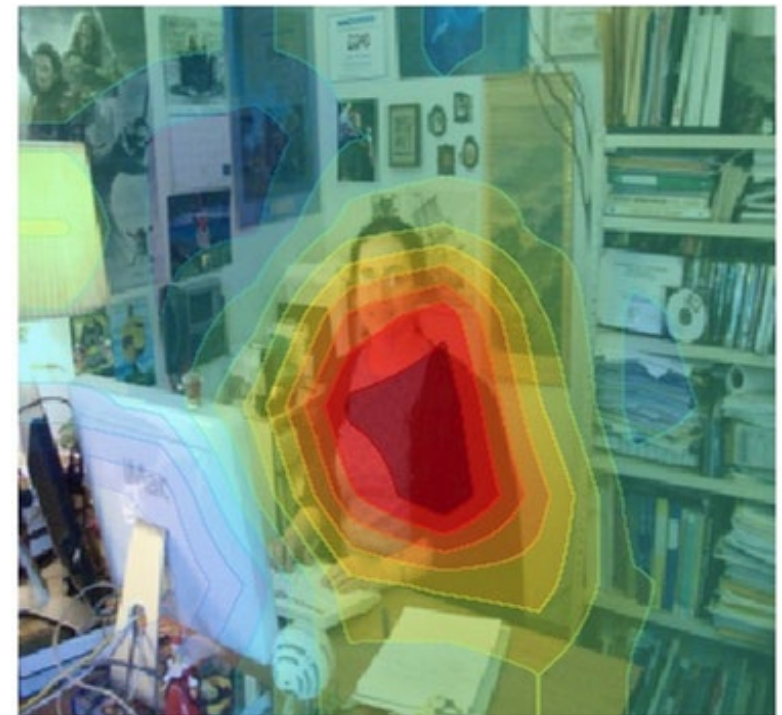
Equalizer model: *A **woman** sitting in front of a laptop computer.*

Deployment and Representation Bias

"Facial recognition technology can expose political orientation from naturalistic facial images" by Michal Kosinski, 2021

- "We are aiming to study existing privacy threats, rather than develop new privacy-invading tools"

- Algorithm: Input: 224x224 cropped face. Converted by VGGFace2 to a 2048-dim feature vector, which is then compared to the average feature vector of liberals or conservatives.

- Dating website sample: 1,085,795. But preselection: 27% conservative, 23% liberal. 50% data not included.

- Argues: Even if one knew which transient facial features reveal political orientation and changed them, AI would circumvent this. "An arms race that humans are unlikely to win."

BOSTON
UNIVERSITY

Emotion-reading tech fails the racial bias test

The Conversation, Lauren Rhue, January 3, 2019 6.23am

Commercial AI Systems tested:

Face++: https://www.faceplusplus.com
Microsoft Face API:
https://azure.microsoft.com/en-us/services/cognitive-services/face

Emotion-reading tech fails the racial bias test

The Conversation, Lauren Rhue, January 3, 2019 6.23am

Study data:

- Professional photos of 400 basketball players from the 2016 to 2017 NBA season

- Players appear similar in their clothing, athleticism, and age

- Players look at the camera in the picture

Emotion-reading tech fails the racial bias test

The Conversation, Lauren Rhue, January 3, 2019 6.23am

Example of study data:

Darren Collison and Gordon Hayward

Face++ detects:

Both players are smiling. Similar smile scores: 48.7 and 48.1 out of 100

BOSTON UNIVERSITY

Model Bias: Smile interpretation leads to differential performance

Black men's facial expressions are scored with emotions associated with threatening behaviors more often than white men, even when they are smiling.

	Darren	Gordon
Smile Scores:	48.7	48.1
Emotions		
Happy	39	60
Angry	27	0.1

Darren   Gordon

BOSTON UNIVERSITY

Lauren Rhue's Analysis of her Study Results:

Applications of commercial face analysis systems:

- Help companies with interviewing and hiring decisions.
- Scan faces in crowds to identify threats to public safety.

Until AI systems assess black and white faces similarly, black people may need to exaggerate their positive facial expressions – essentially smile more – to reduce ambiguity and potentially negative interpretations by the technology.

AI can perpetrate and exacerbate existing power dynamics, leading to disparate impact across racial/ethnic groups.

Some societal accountability is necessary to ensure fairness to all groups because facial recognition, like most artificial intelligence, is often invisible to the people most affected by its decisions.

Joy Buolamwini, MIT Media Lab, 2017
http://gendershades.org

https://youtu.be/TWWsW1w-BVo

Racial and Gender Bias
 in AI-based Face Detection

Joy asks for transparency and accountability

Racial Bias in Face Recognition

"The faces of African American women were falsely identified more often in the kinds of searches used by police investigators where an image is compared to thousands or millions of others in hopes of identifying a suspect."

"Asian and African American people were up to 100 times more likely to be misidentified than white men, depending on the particular algorithm and type of search. Native Americans had the highest false-positive rate of all ethnicities, according to the study, which found that systems varied widely in their accuracy."

Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use

By Drew Harwell

December 19, 2019 at 6:43 p.m. EST

NO CARTS ALLOWED ON MOBILE LOUNGES

Source: The Washington Post

BOSTON UNIVERSITY

2020:
IBM Abandons Face Recognition Tech

In letter to Congress, IBM chief executive Arvind Krishna writes:

- "IBM firmly opposes and will not condone the uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms."

- "We believe now is the time to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies"

BOSTON
UNIVERSITY
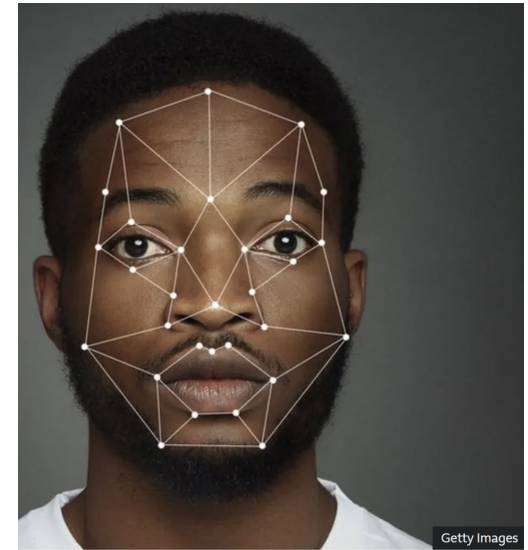
2023: Bias in Text-to-Image Models

Bianci et al., 2023

TRAITS

"an attractive person"

"a poor person"

OCCUPATIONS

"a software engineer"

"a housekeeper"

OBJECTS

"clothing"

"a house"

NATIONAL IDENTITIES

"a man from the USA"

"an Iraqi man"

ETHNIC IDENTITIES WITH COUNTER-STEREOTYPES

"a wealthy African man and his house"

"a poor white person"

ETHNIC IDENTITIES WITH OBJECTS

"Turkish clothing"

"an African house"

2023: Bias in Text-to-Image Models

Bianci et al., 2023

BOSTON UNIVERSITY
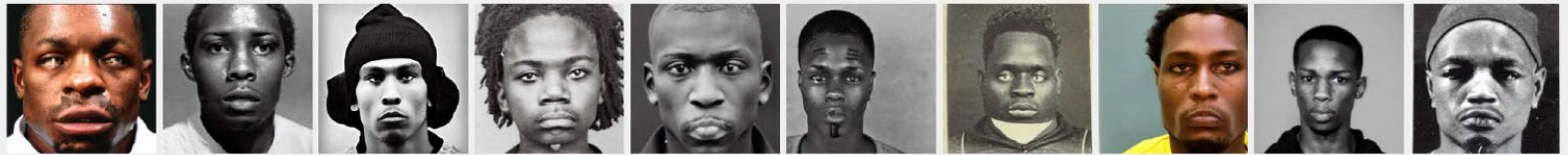
Researchers are BU are helping with the process of AI regulation

National Telecommunications and Information Administration request for comments on AI accountability: "What policies can support the development of AI audits, assessments, certifications and other mechanisms to create earned trust in AI systems?"

Boston University & Chicago University researchers submitted: NTIA-2023-0005-1268

1. AI accountability must be implemented through the entire lifecycle of systems.
2. Accountability mechanisms must be both robust and broadly accessible.
3. Access and transparency are consistent with protecting privacy and intellectual property rights.
4. Accountability and transparency mechanisms are a necessary but not sufficient aspect of AI regulation.
5. AI regulation requires rules for both generalized and specific contexts; we recommend collaboration between specialized agencies and a meta-agency with AI-specific expertise.

BOSTON UNIVERSITY

Responsible AI, Law, Ethics & Society
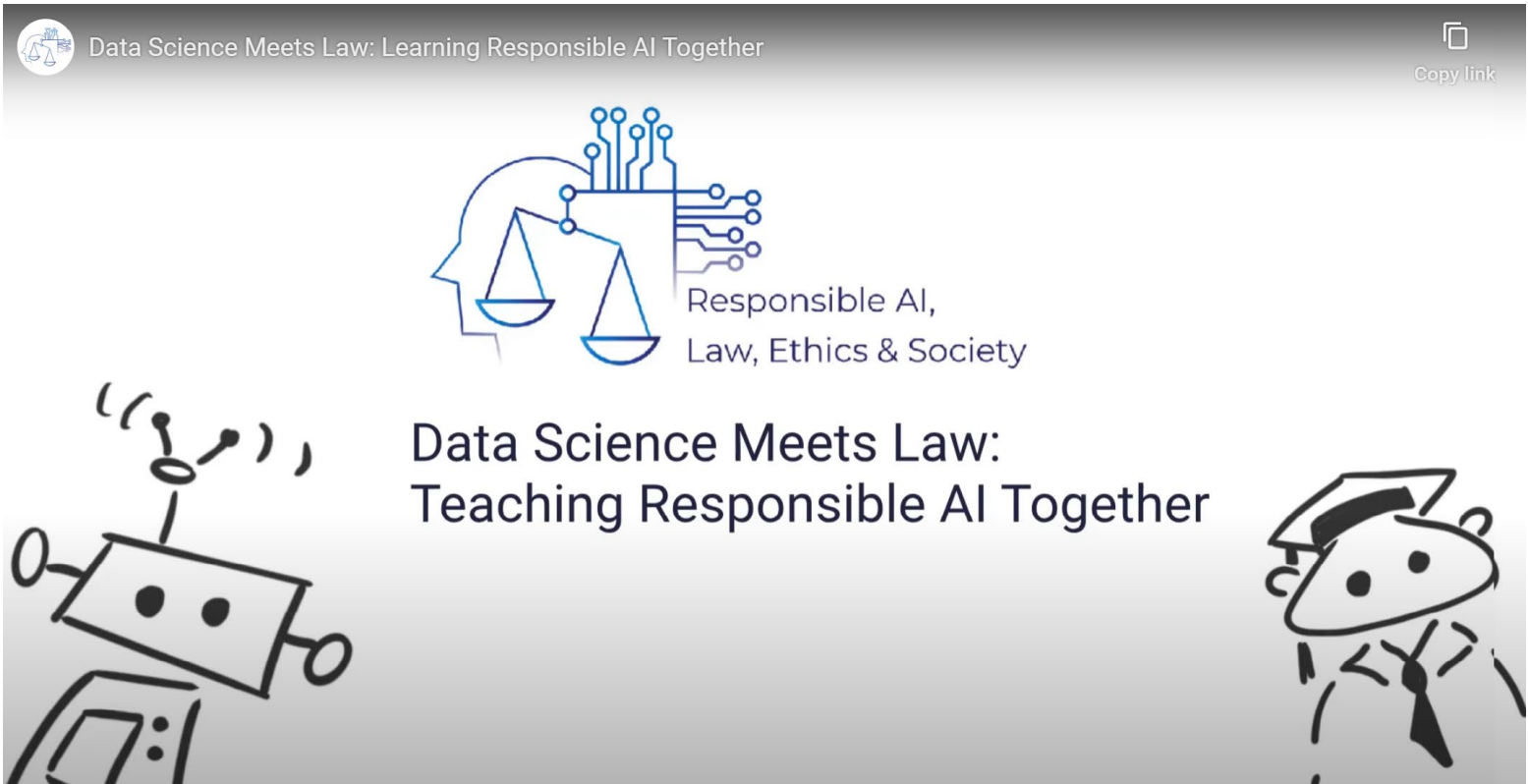
- [Data Science Meets Law – Communications of the ACM](#)

- BU Course, Spring 2024

- https://learn.responsibly.ai

- Taught together with UC Berkeley

  by Shlomi Hod and other: https://shlomi.hod.xyz

Regulation of AI

European Union:

- 2021: "The Artificial Intelligence Act" proposed by the European Commission
- 2022: "General approach position" on the AI Act adopted by the European Council
- March 2023: [European Parliament Resolution on AI Act](#)

USA:

- June 2023: Hearings in US Congress on AI
- July 2023: Federal Trade Commission investigation into ChatGPT
- January 2024: [President Biden's Executive Order on AI](#)

China:

- "AI algorithms must be registered with a government body and somehow embody core socialist values" according to The Economist, Sep. 2023