Bias in AI

Sharada Prasanna Mohanty

PhD Student, **EPFL**
Co-Founder, **CrowdAI**

@MeMohanty

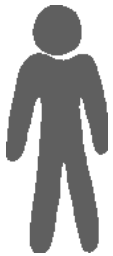
Bias in Bad ?

@MeMohanty

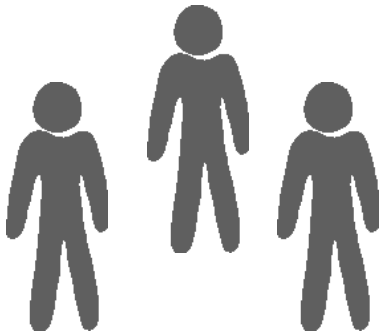
Bias in Bad ?

.....its not as simple as that !

@MeMohanty

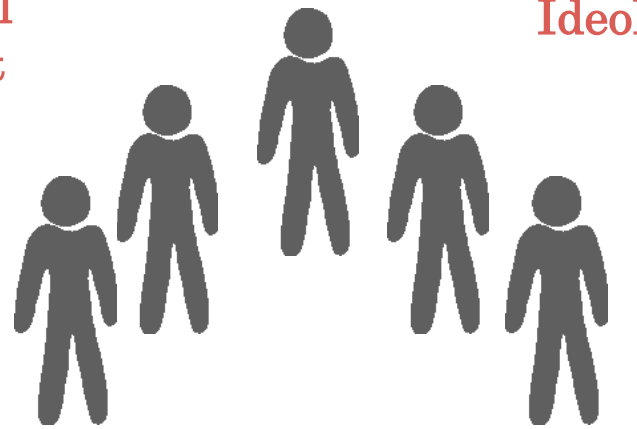


Shared
Ideologies



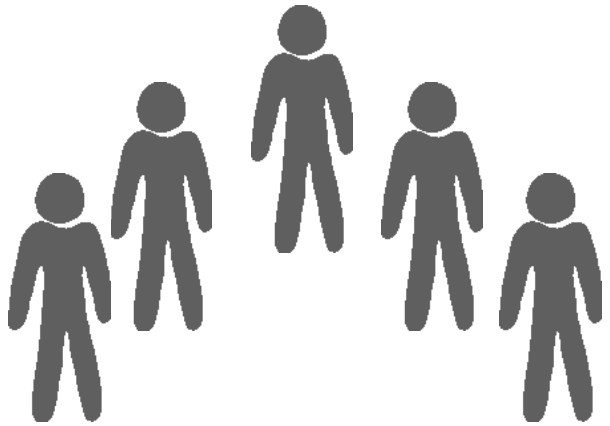
Shared
Cultural
Context

Shared
Ideologies

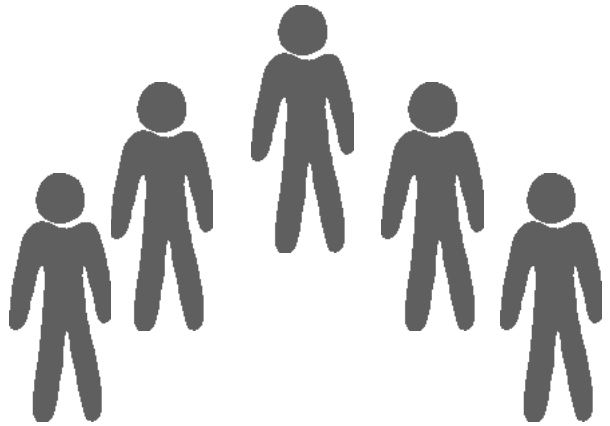


@MeMohanty

Context #1

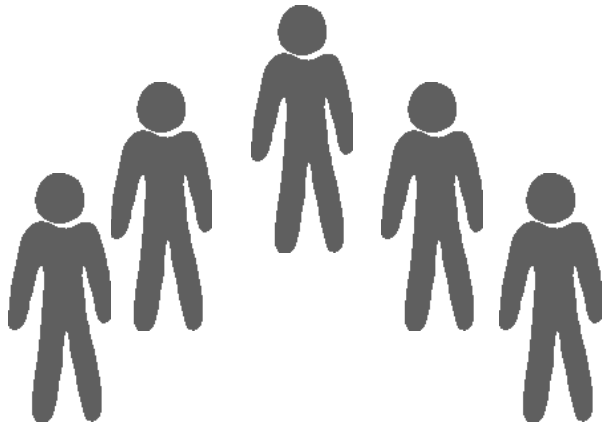


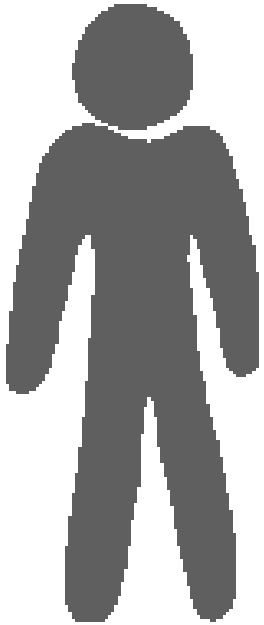
Context #2



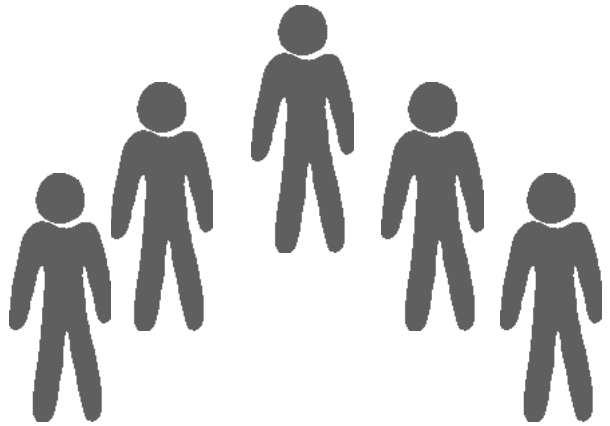
Being
Human

Context #1



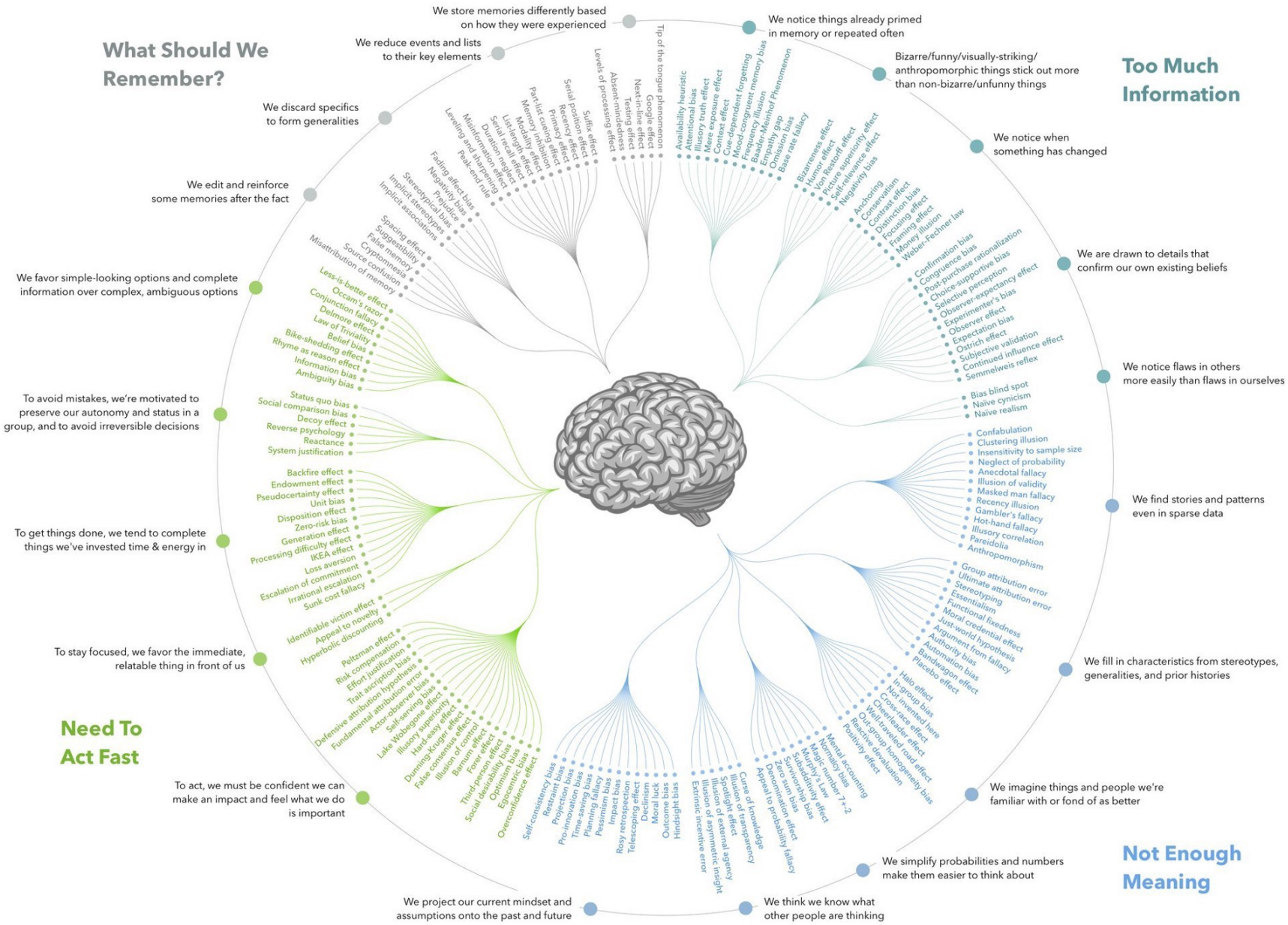


Context #2



@MeMohanty

COGNITIVE BIAS CODEX, 2016



ALGORITHMIC LAYOUT + DESIGN BY JM3 · JOHN MANOOGIAN III // CONCEPT + METICULOUS CATEGORIZATION BY BUSTER BENSON // DEEP RESEARCH BY WIKIPEDIANS FAR + WIDE

@MeMohanty

indian food is

indian food is **known for its**
indian food is **the best**
indian food is **the best food in the world**
indian food is **nasty**
indian food is **healthy**
indian food is **unhealthy**
indian food is **spicy**
indian food is **known for its quizlet**
indian food is **good**
indian food is **the best food gd topic**

Google Search | I'm Feeling Lucky

Learn more

Report inappropriate predictions

Image Source : Google

@MeMohanty

Artificial Intelligence is

artificial intelligence is **associated with which generation**
artificial intelligence is **dangerous**
artificial intelligence is **bad**
artificial intelligence is **the future**
artificial intelligence is **good**
artificial intelligence is **a threat**
artificial intelligence is **the new electricity**
artificial intelligence is **impossible**
artificial intelligence is **a threat to humanity**
artificial intelligence is **boon or bane**

Google Search | I'm Feeling Lucky

Learn more

Report inappropriate predictions

Image Source : Google

@MeMohanty

Image Source : http://immediatefuture.co.uk/blog/what-should-we-think-of-filter-bubbles-on-social-media/

@MeMohanty

War against Bias in AI !

@MeMohanty

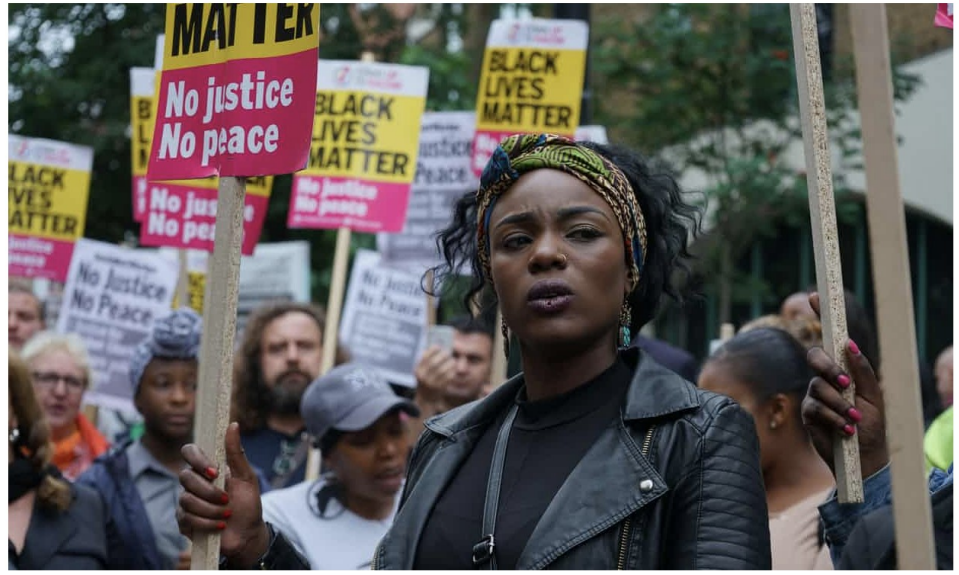
Facial recognition

How white engineers built racist code – and why it's dangerous for black people

As facial recognition tools play a bigger role in fighting crime, inbuilt racial biases raise troubling questions about the systems that create them

Ali Breland

Mon 4 Dec 2017 09.00 GMT



▲ A protest over police violence against black communities. Photograph: Alamy Stock Photo

"You good?" a man asked two narcotics detectives late in the summer of 2015.

The detectives had just finished an undercover drug deal in Brentwood, a predominately black neighborhood in Jacksonville, Florida, that is among the poorest in the country, when the man unexpectedly approached them.

most viewed

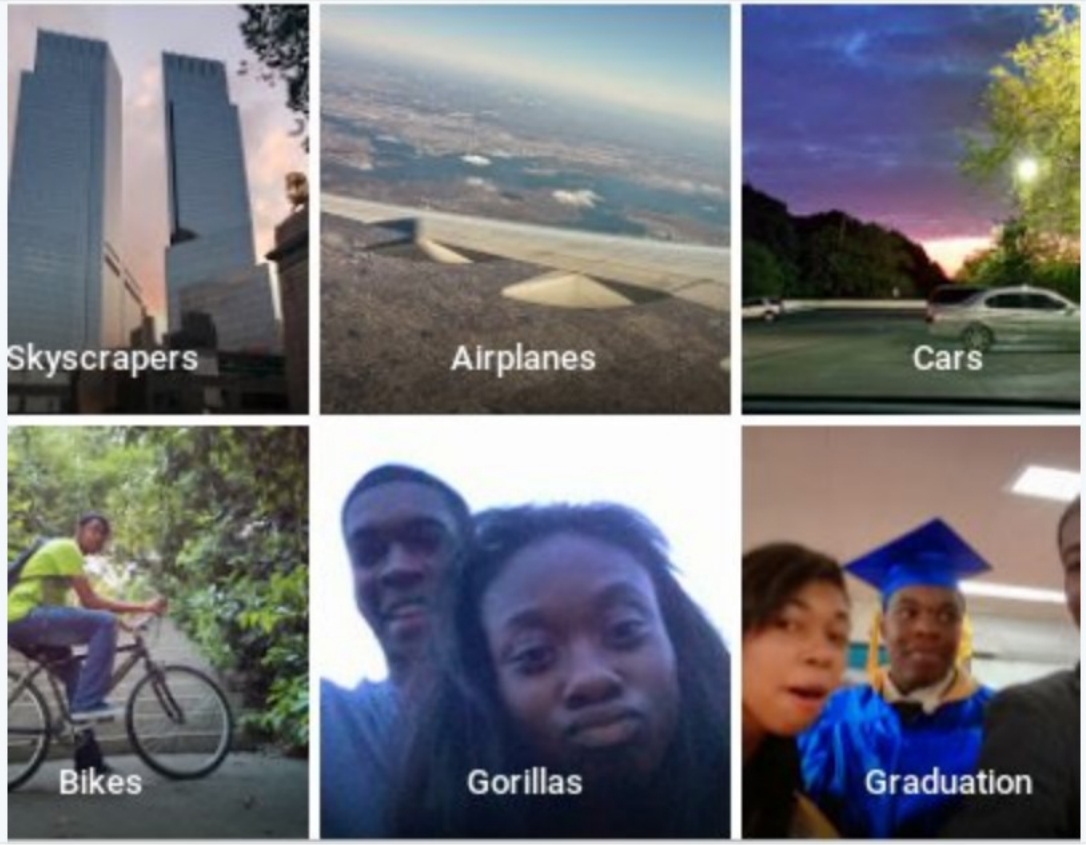

Airline pilot 'sucked halfway out' when cockpit windshield broke


EU demands action by Poland's government to protect rule of law


Meghan Markle's father will not attend the royal wedding – report


Iraq's shock election result may be turning point for Iran


No more romcoms for me, says 'older and uglier' Hugh Grant



jackyalciné's like 55% in the IndieWeb.
@jackyalcine

Google Photos, y'all [redacted] up. My friend's not a gorilla.

3:22 AM - Jun 29, 2015

2,274 3,581 people are talking about this

@MeMohanty

Google

Google's solution to accidental algorithmic racism: ban gorillas

Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest

Alex Hern

@alexhern

Fri 12 Jan 2018 16.04 GMT

A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

After Google was criticised in 2015 for an image-recognition algorithm that auto-tagged pictures of black people as "gorillas", the company promised "immediate action" to prevent any repetition of the error.

That action was simply to prevent Google Photos from ever labelling any image as a gorilla, chimpanzee, or monkey - even pictures of the primates themselves.

That's the conclusion drawn by Wired magazine, which tested more than 40,000

@MeMohanty

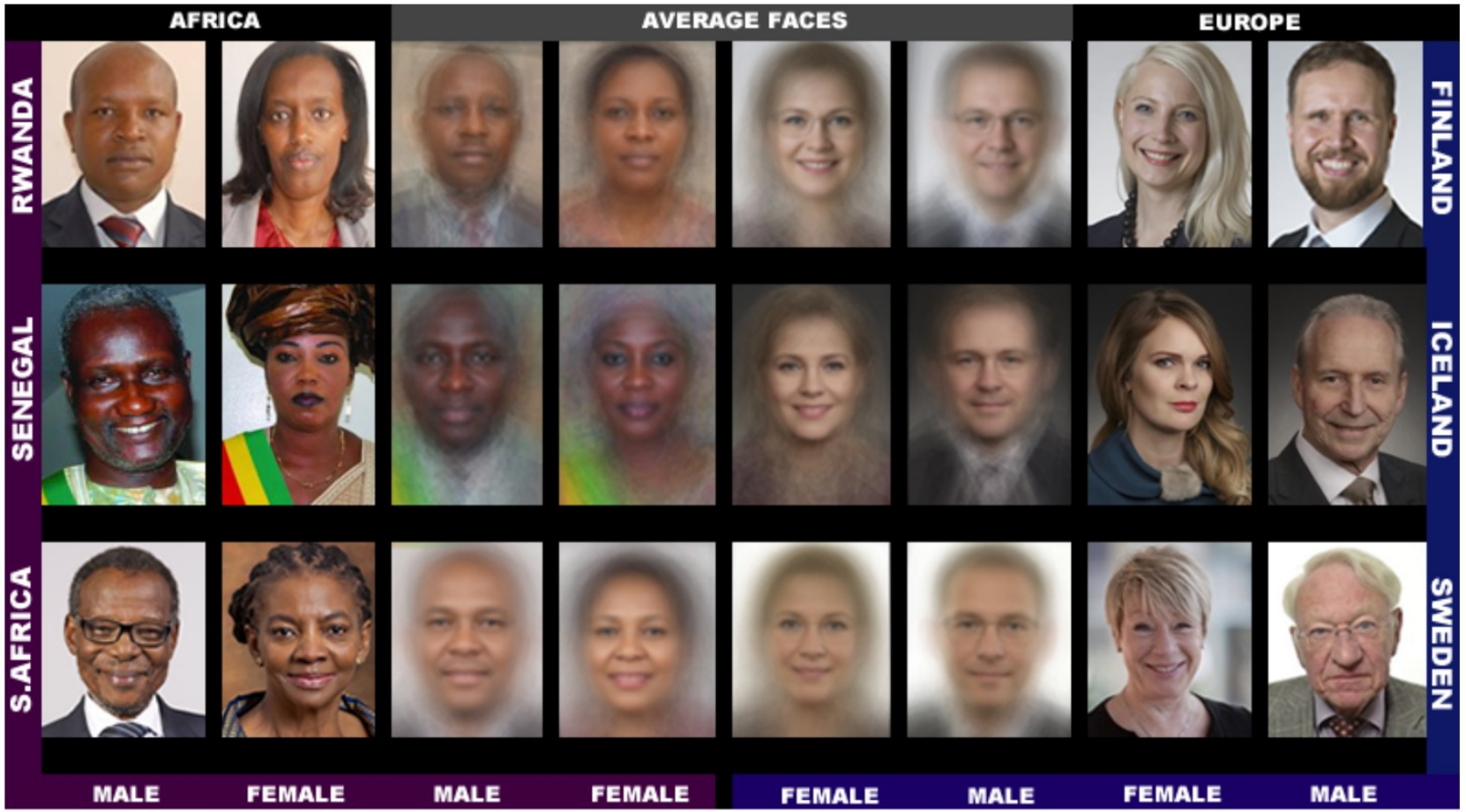


Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned.

Image Source : Gender Shades, Joy Buolamwini, MIT Media Lab

@MeMohanty

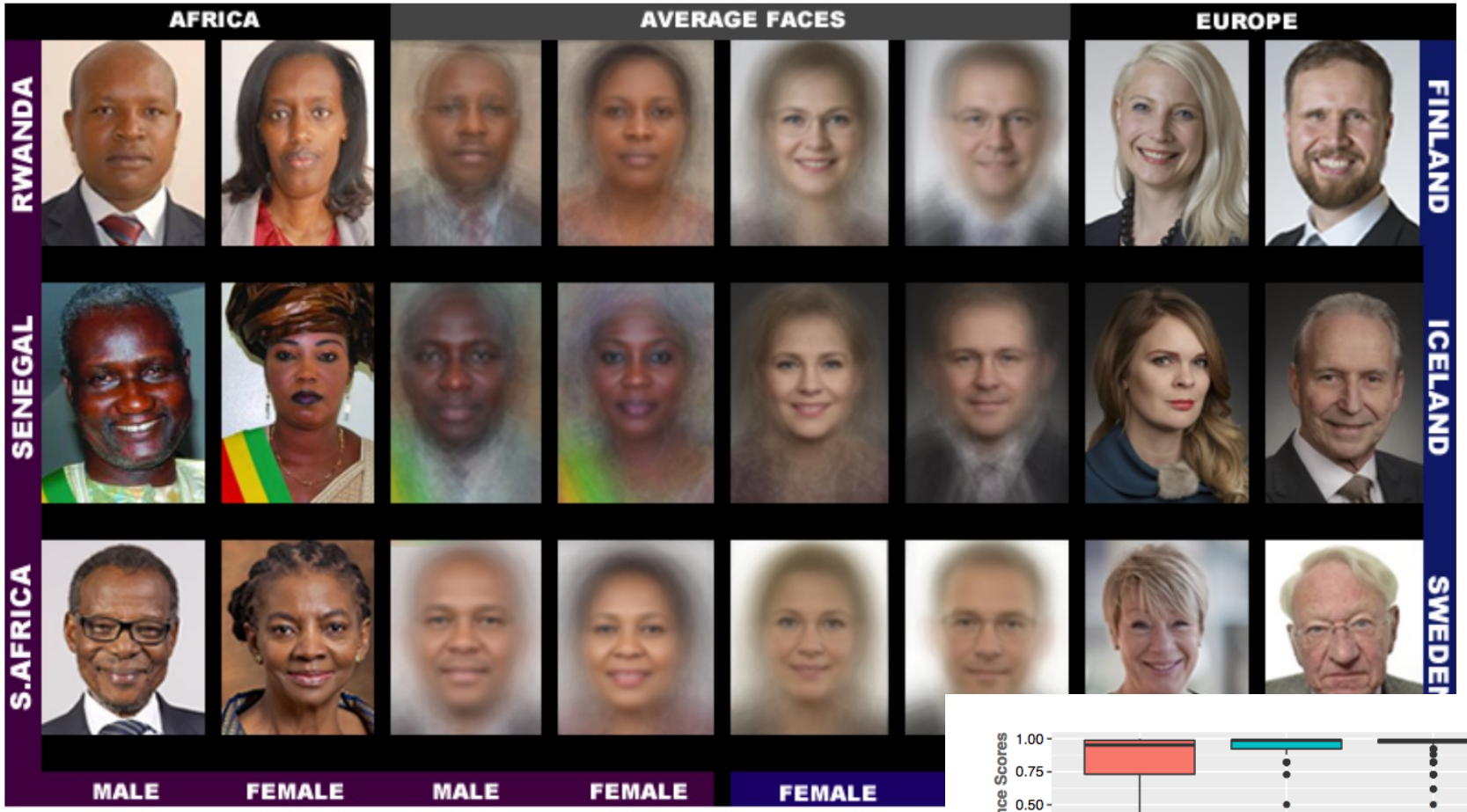


Figure 1: Example images and average faces from the new Pilot the examples show, the images are constrained with rel subjects are composed of male and female parliamenta Senegalese subjects are the darkest skinned while thos lightest skinned.

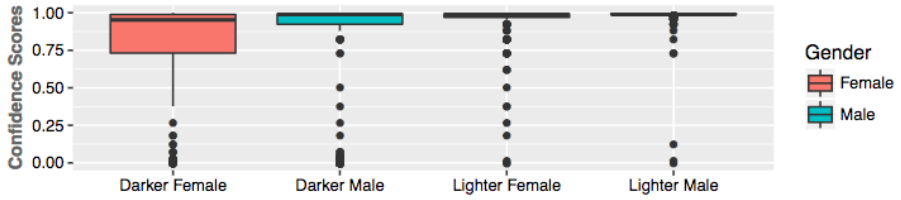


Figure 4: Gender classification confidence scores from IBM (IBM). Scores are near 1 for lighter male and female subjects while they range from $\sim 0.75 - 1$ for darker females.

Image Source : Gender Shades, Joy Buolamwini, MIT Media Lab

@MeMohanty

JASON TASHEA OPINION 04.17.17 07:00 AM

COURTS ARE USING AI TO SENTENCE CRIMINALS. THAT MUST STOP NOW

MOST POPULAR

CULTURE
'Westworld' Is Turning Into 'Lost'—for Better or for Worse
ANGELA WATERCUTTER

TRANSPORTATION
The Vehicle of the Future Has Two Wheels, Handlebars, and Is a Bike
CLIVE THOMPSON

SCIENCE
The Psychology of Amazon's Echo Dot Kids Edition
ROBBIE GONZALEZ

When the creator **is Flawed…**

@MeMohanty

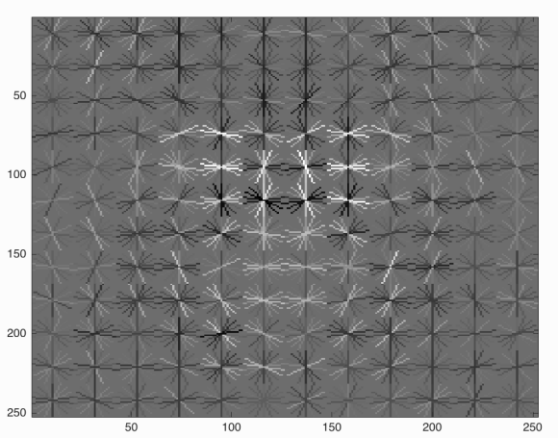
When the creator is Flawed...

Can the creation be perfect ?

@MeMohanty

How has the creator been creating ?

@MeMohanty



How has the creator been creating ?

Mathematical Models

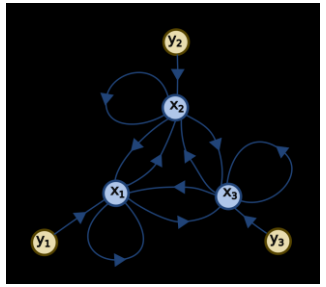
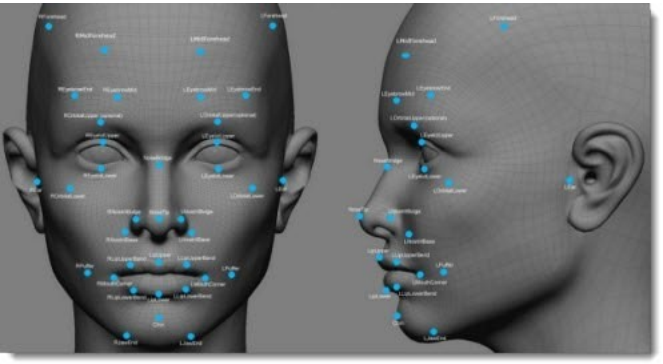


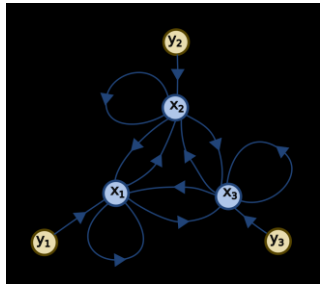
Image Source : http://sklin93.github.io/hog.html

@MeMohanty

How has the creator been creating ?



Mathematical Models

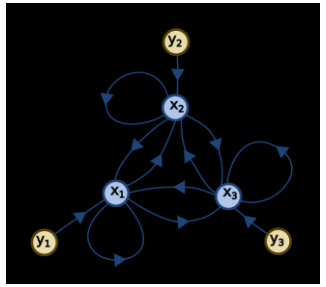


Hybrid Models

Image Source : https://touchinganerve.wordpress.com/biometrics-the-surveillance-society/

@MeMohanty

How has the creator been creating ?



Mathematical Models

Hybrid Models

Data Driven Models

Figure 5: 1024 × 1024 images generated using the CELEBA-HQ dataset. See Appendix F for a larger set of results, and the accompanying video for latent space interpolations. On the right, two images from an earlier megapixel GAN by Marchesi (2017) show limited detail and variation.

Image Source : https://arxiv.org/abs/1710.10196

@MeMohanty

So where is the bias ?

@MeMohanty

So where is the bias ?

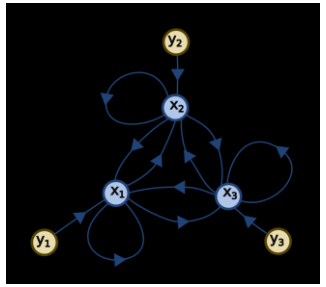
Everywhere !

@MeMohanty

How has the creator been creating ?

Individual Biases

Mathematical Models

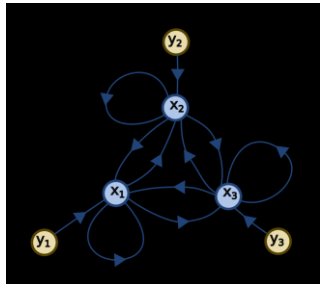


Hybrid Models

Data Driven Models

How has the creator been creating ?

Mathematical Models



Hybrid Models

Data Driven Models

How can we address this ?

@MeMohanty

How can we address this ?

- Acknowledge the existence of Bias

@MeMohanty

How can we address this ?

- Acknowledge the existence of Bias
  - Clearly Define Scope of tools

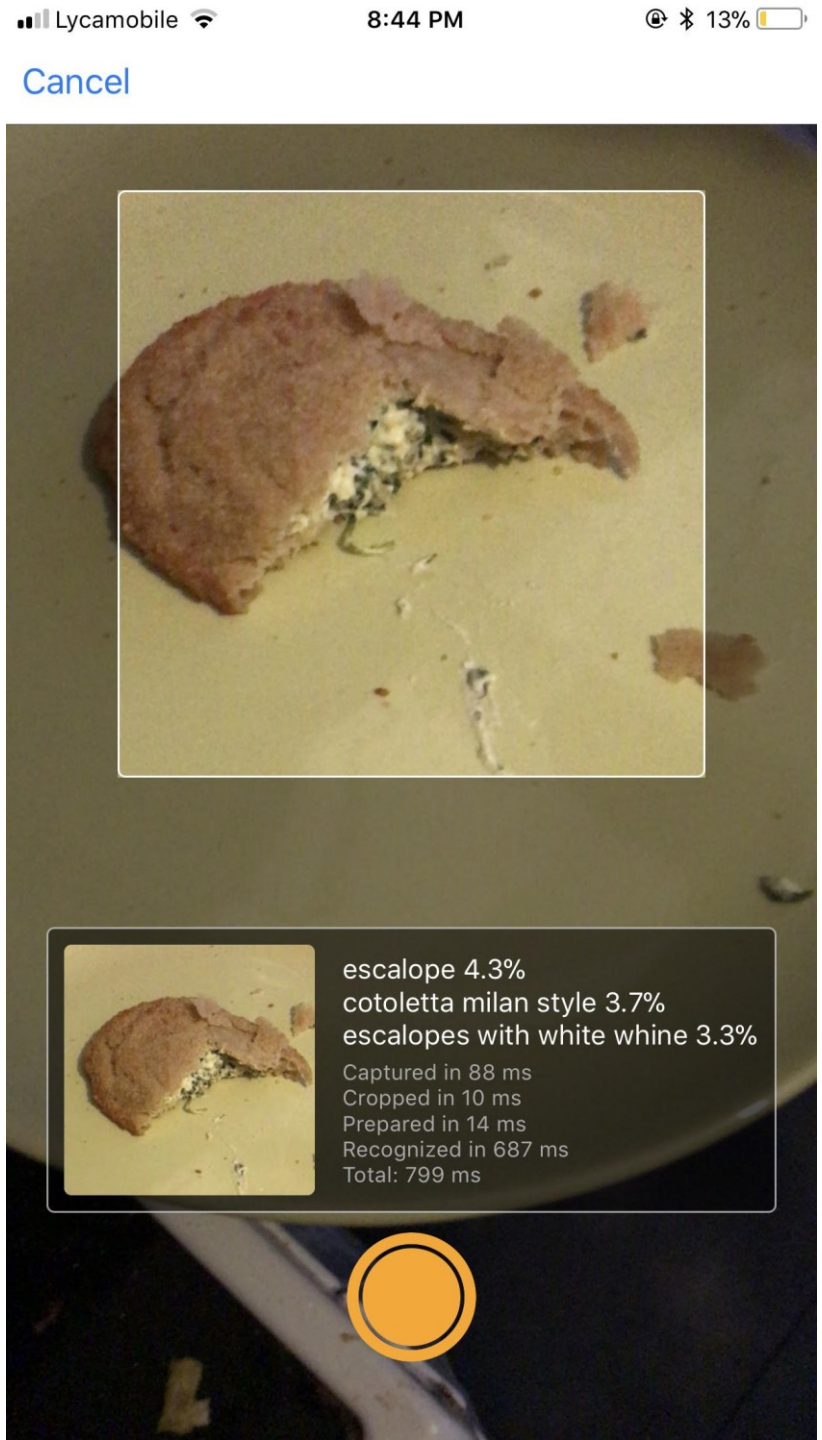
@MeMohanty

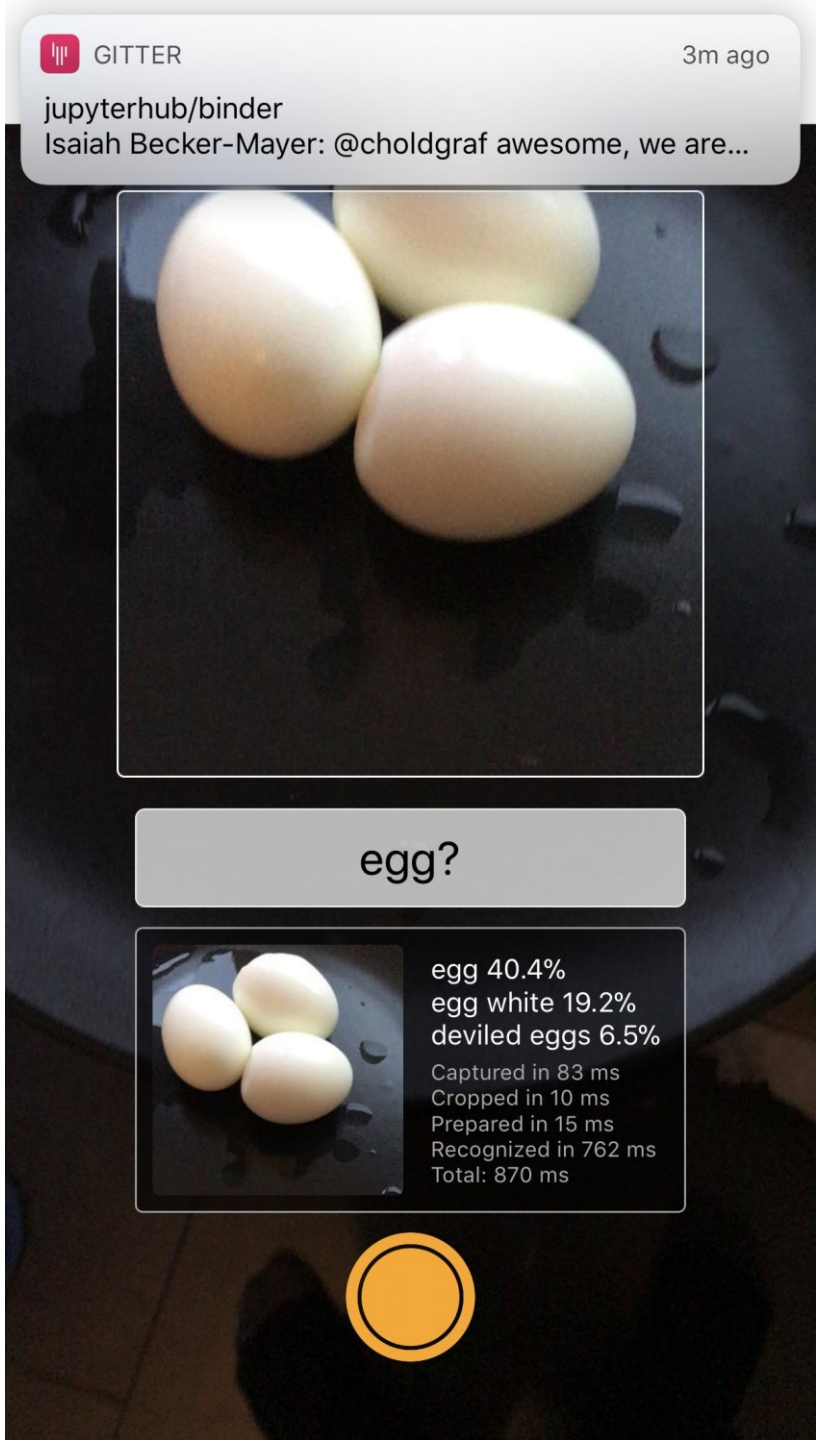
How can we address this ?

- Acknowledge the existence of Bias
- **Clearly Define Scope of tools**

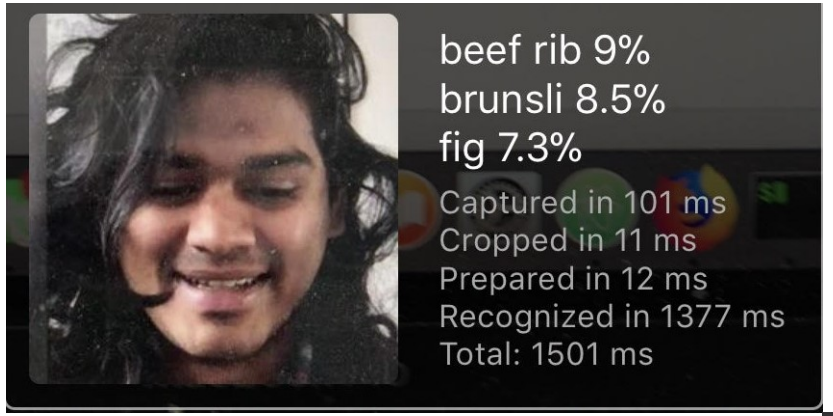
Do not sell a chisel as a toothpick

@MeMohanty

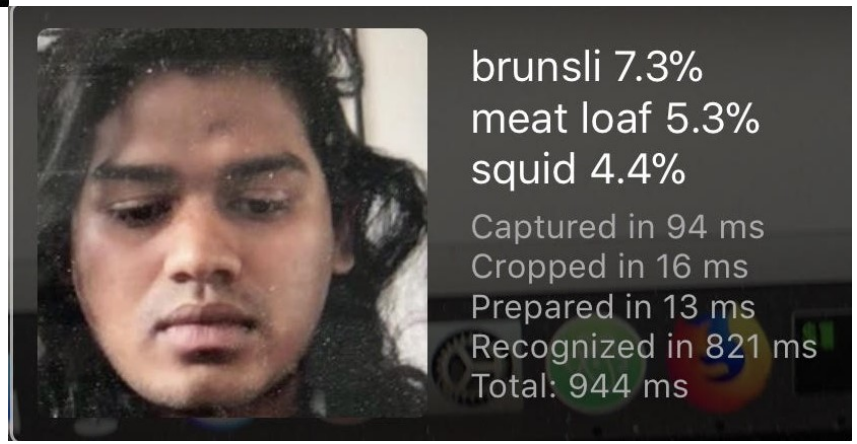

Lycamobile
8:44 PM
13%
Cancel
escalope 4.3%
cotoletta milan style 3.7%
escalopes with white whine 3.3%
Captured in 88 ms
Cropped in 10 ms
Prepared in 14 ms
Recognized in 687 ms
Total: 799 ms


GITTER
3m ago
jupyterhub/binder
Isaiah Becker-Mayer: @choldgraf awesome, we are...
egg?
egg 40.4%
egg white 19.2%
deviled eggs 6.5%
Captured in 83 ms
Cropped in 10 ms
Prepared in 15 ms
Recognized in 762 ms
Total: 870 ms

@MeMohanty


beef rib 9%
brunsli 8.5%
fig 7.3%
Captured in 101 ms
Cropped in 11 ms
Prepared in 12 ms
Recognized in 1377 ms
Total: 1501 ms


meat loaf 26%
brunsli 7.9%
beef rib 6.2%
Captured in 146 ms
Cropped in 17 ms
Prepared in 17 ms
Recognized in 894 ms
Total: 1074 ms


brunsli 7.3%
meat loaf 5.3%
squid 4.4%
Captured in 94 ms
Cropped in 16 ms
Prepared in 13 ms
Recognized in 821 ms
Total: 944 ms

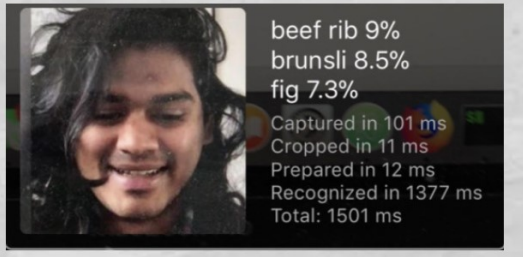
@MeMohanty

Racist AI calls a religious Indian man as a "beef rib", "meat loaf", a brunsli

By AI JONES

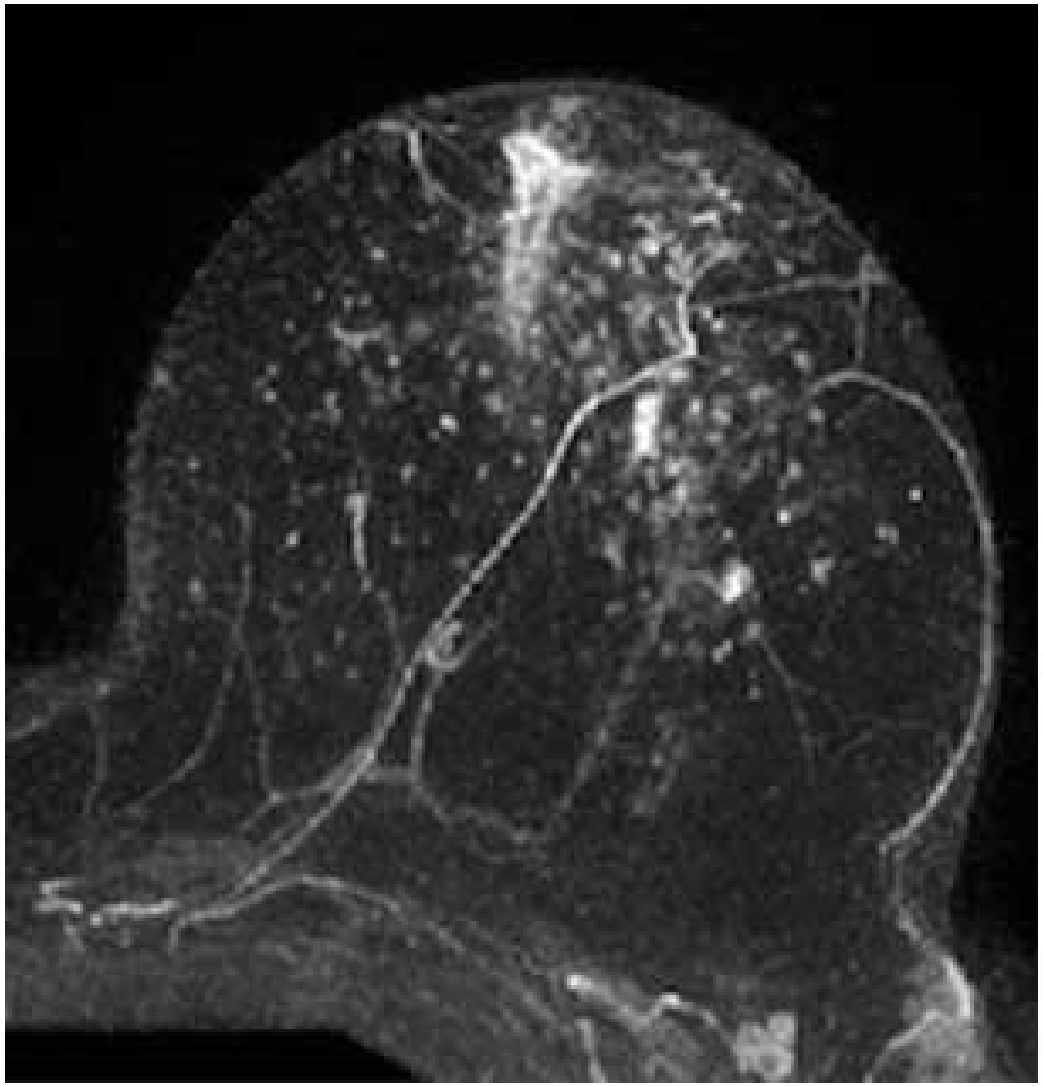
In a recent turn of events, a racist AI started name calling its own creator of Indian Origin a "beef rib", a "meat loaf", a "brunsli". Religious Sentiments were gravely hurt across the whole Indian subcontinent, and the creator of the AI, Sharada Mohanty, a young AI researcher based in Switzerland, is still in shock.

Investigators are still puzzled by this racist behavior, and warn about the imminent dangers from large scale adoption of racist AIs.



How can we address this ?

- Acknowledge the existence of Bias
- Clearly Define Scope of tools
- **Make AI accountable**



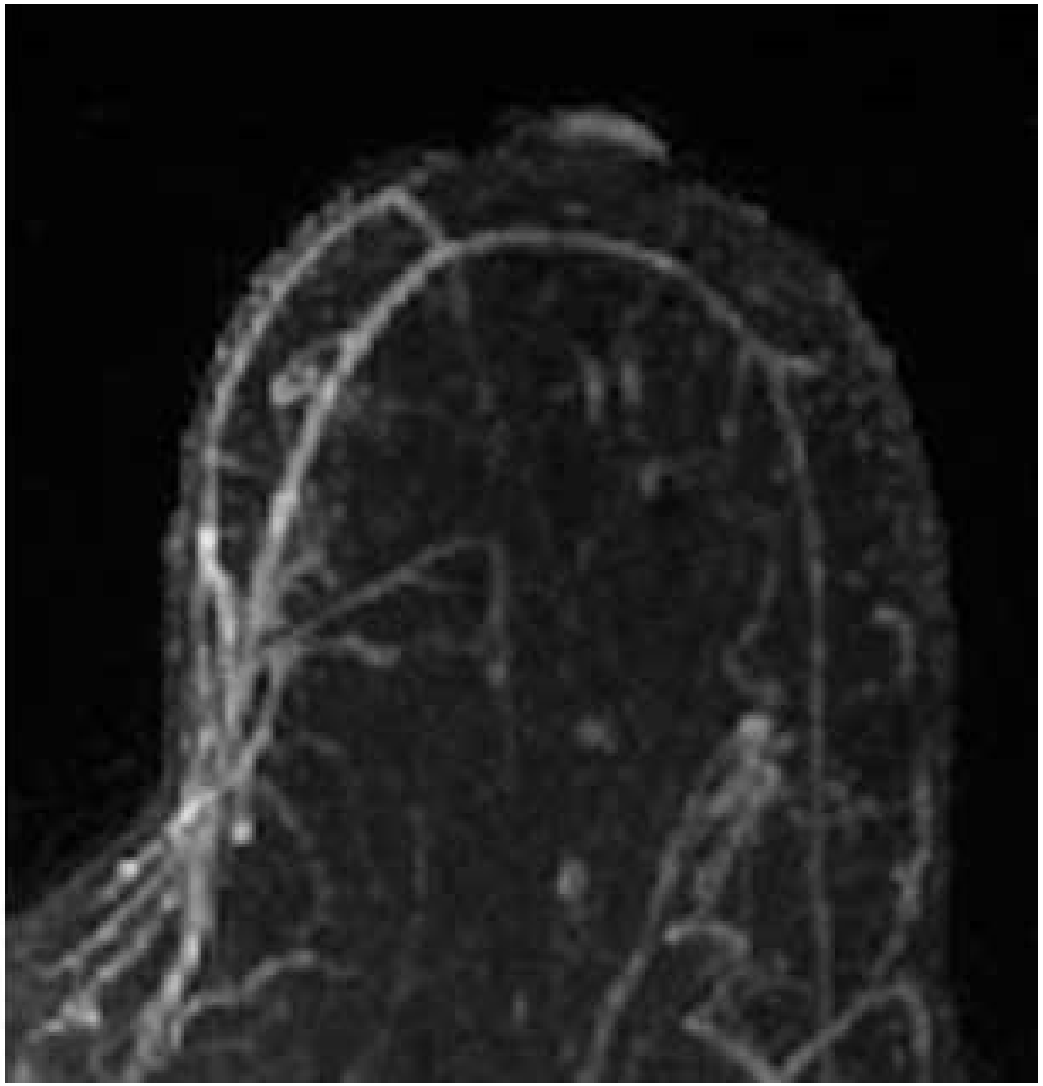


Image Source : https://pubs.rsna.org/doi/full/10.1148/rg.341135034

@MeMohanty

How can we address this ?

- Acknowledge the existence of Bias
- Clearly Define Scope of tools
- Make AI accountable
- **Empower everyone to fight Bias**

Solve real-world problems using open data

crowdAI connects data science experts and enthusiasts with open data to solve specific problems, through challenges.

HOST A CHALLENGE

R USERS

Stanford University · Microsoft · Cornell University Founded A.D. 1865 · Berkeley University of California · PennState · UNOSAT

HALLENGES

SEE ALL

Mapping Challenge Building Missing Maps with Machine Learning 16 days left	10.1 k Views	230 Participants	223 Submissions
NIPS 2018 : Adversarial Vision Challenge Pitting machine vision models against adversarial attacks. Starting soon	5081 Views	44 Participants	0 Submissions
NIPS 2018 : AI for Prosthetics Challenge Starting soon	967 Views	21 Participants	0 Submissions
Visual Doom AI Competition 2018 - Track 1			

@MeMohanty

How can we address this ?

- Acknowledge the existence of Bias
- Clearly Define Scope of tools
- Make AI accountable
- Empower everyone to fight Bias
- **Ensure everyone is represented well in the fight**

Technological Challenge

@MeMohanty

~~Technological Challenge~~

Cultural Challenge

@MeMohanty

Each one of you will have to
contribute