

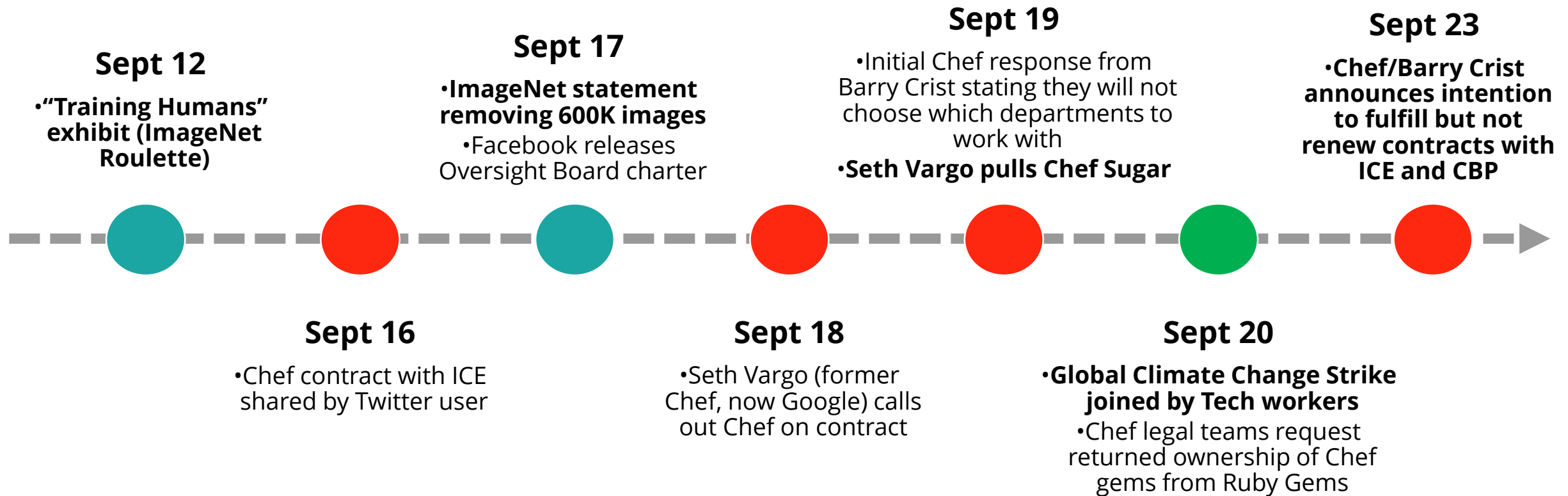


Who Watches the Watchers?

#ethics in a #machinelearning world

Jason St-Cyr
@StCyrThoughts







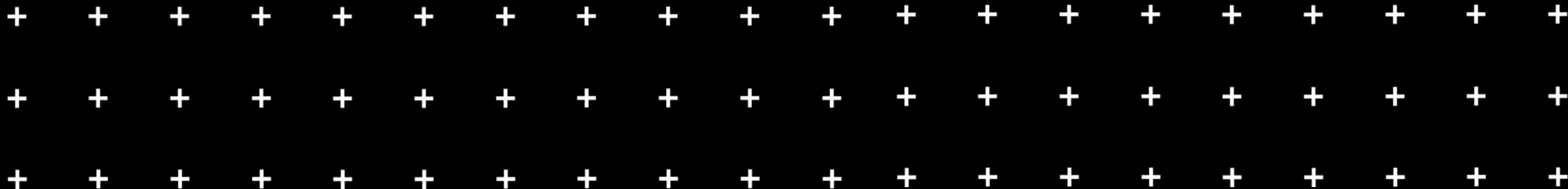
Virginia Dignum

@vdignum

Replying to [@DorotheaBaur](#) [@StCyrThoughts](#) and 8 others

May be the difference is between doing the right thing
(ethics) and doing the thing right (Responsibility)

5:47 PM · Apr 23, 2019 · [Twitter for Android](#)



— All my instincts they return...

@StCyrThoughts

We can take the time to think...

"Is this the right thing to do?"



Photo by Alex Eylar (<https://www.flickr.com/photos/hoyvinmayvin/>)





Checking restaurant hours

@StCyrThoughts



ON THE TOPIC OF AI:

I just watched the @Google #duplex video. Absolutely amazing. As AI grows, it's not the machines that I'm afraid of, **it's the humans that control the machines that frighten me.** AI will be used by the super rich to manipulate, deceive, and exploit us.

8:50 AM – 10 May 2018

@Peesha_Deel, *Swirling Archer*

realizing. The obvious question soon followed: Should AI software that's smart enough to trick humans be forced to disclose itself. Google executives

Mark Bergen, *Bloomberg*

Google Assistant making calls pretending to be human not only without disclosing that it's a bot, but adding "ummm" and "aaah" to deceive the human on the other end with the room cheering it... horrifying. Silicon Valley is ethically lost, rudderless and has not learned a thing.

8:12 AM – 9 May 2018

@Zeynep, *Zeynep Tufekci*
UNCSILS professor & NYTimes writer

With no exceptions so far, the sense of these reactions has confirmed what I suspected – that people are just fine with talking to automated systems **so long as they are aware of the fact that they are not talking to another person**. They react viscerally and negatively to the concept of machine-based systems that have the effect (whether intended or not) of fooling them into believing that a human is at the other end of the line. To use the vernacular: **“Don’t try to con me, bro!”**

Lauren Weinstein, *People for Internet Responsibility*

**YOUR SCIENTISTS WERE SO PREOCCUPIED WITH
WHETHER OR NOT THEY COULD**

**THAT THEY DIDN'T STOP TO
THINK IF THEY SHOULD**

Who asks the question?

@StCyrThoughts



Artificial Intelligence “around the world”

@StCyrThoughts



-  **Amazon, Google**
Top-ranked AI groups
-  **Nvidia**
Part of the growing trend of AI/ML research in Canada
-  **Microsoft, Apple, Intel**
Major players ranked behind leaders
-  **IBM, Facebook**
North Eastern US-based, growing globally

Digital Marketing and Bias



Impact of bias: Exclusion

@StCyrThoughts



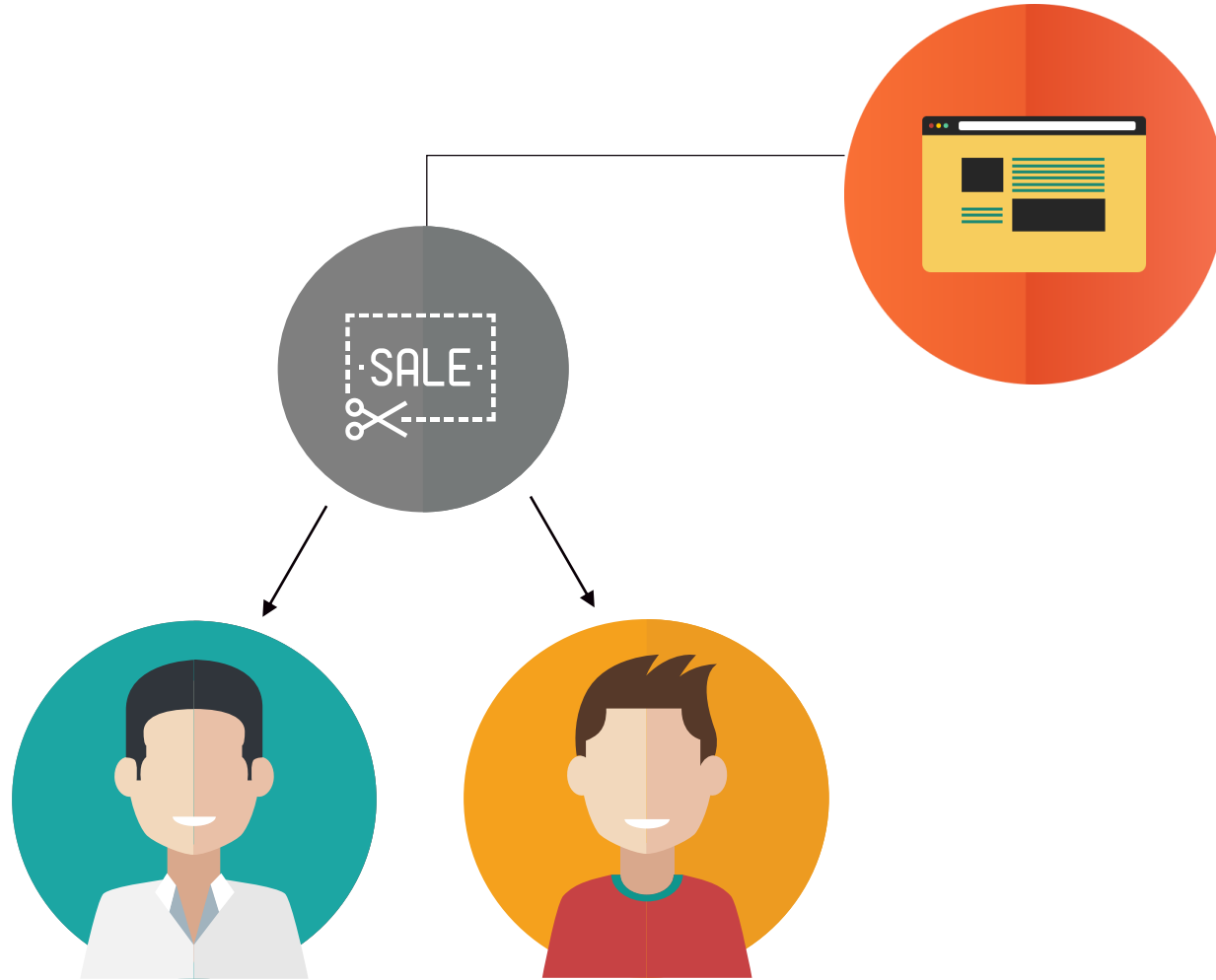
Impact of bias: Exclusion

@StCyrThoughts



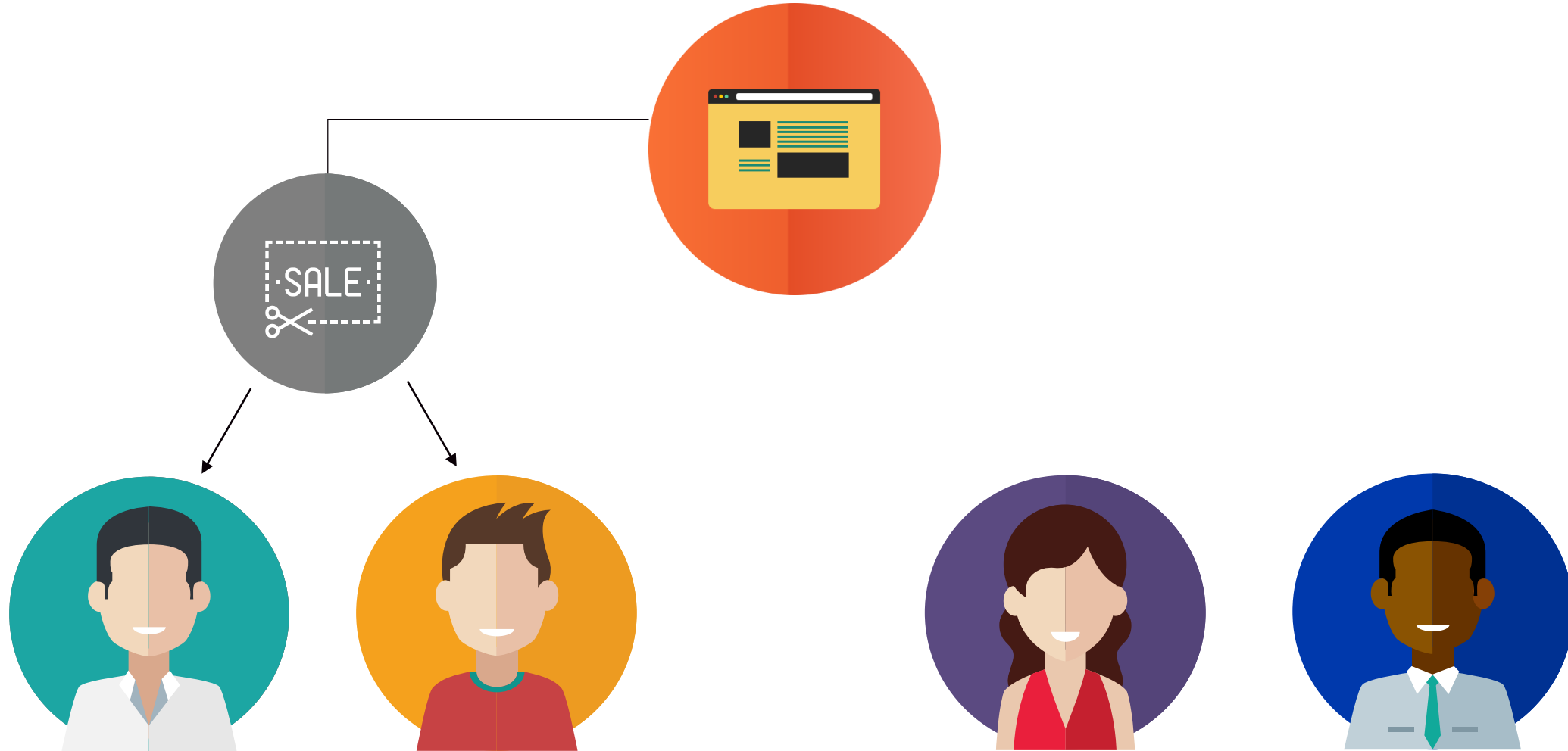
Impact of bias: Exclusion

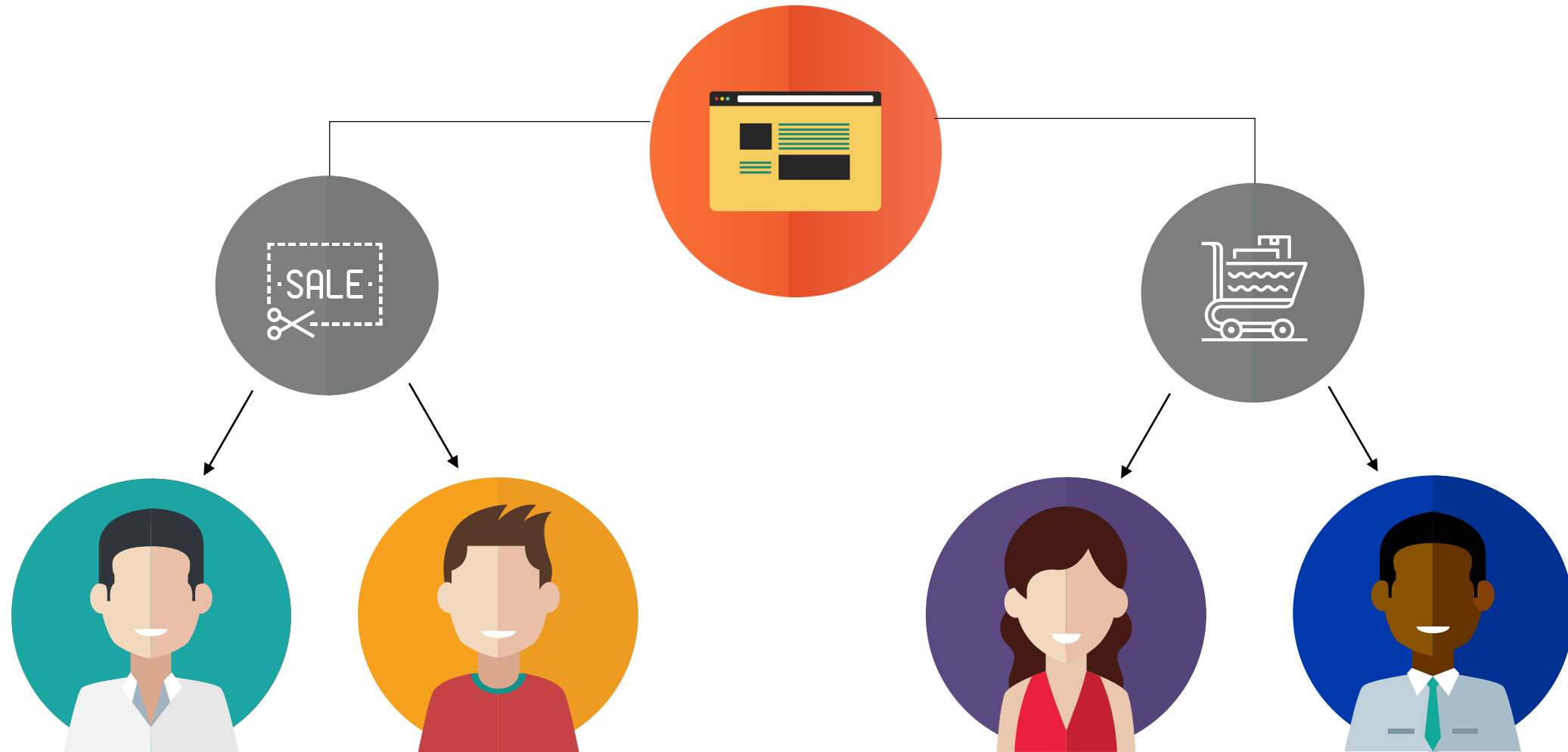
@StCyrThoughts

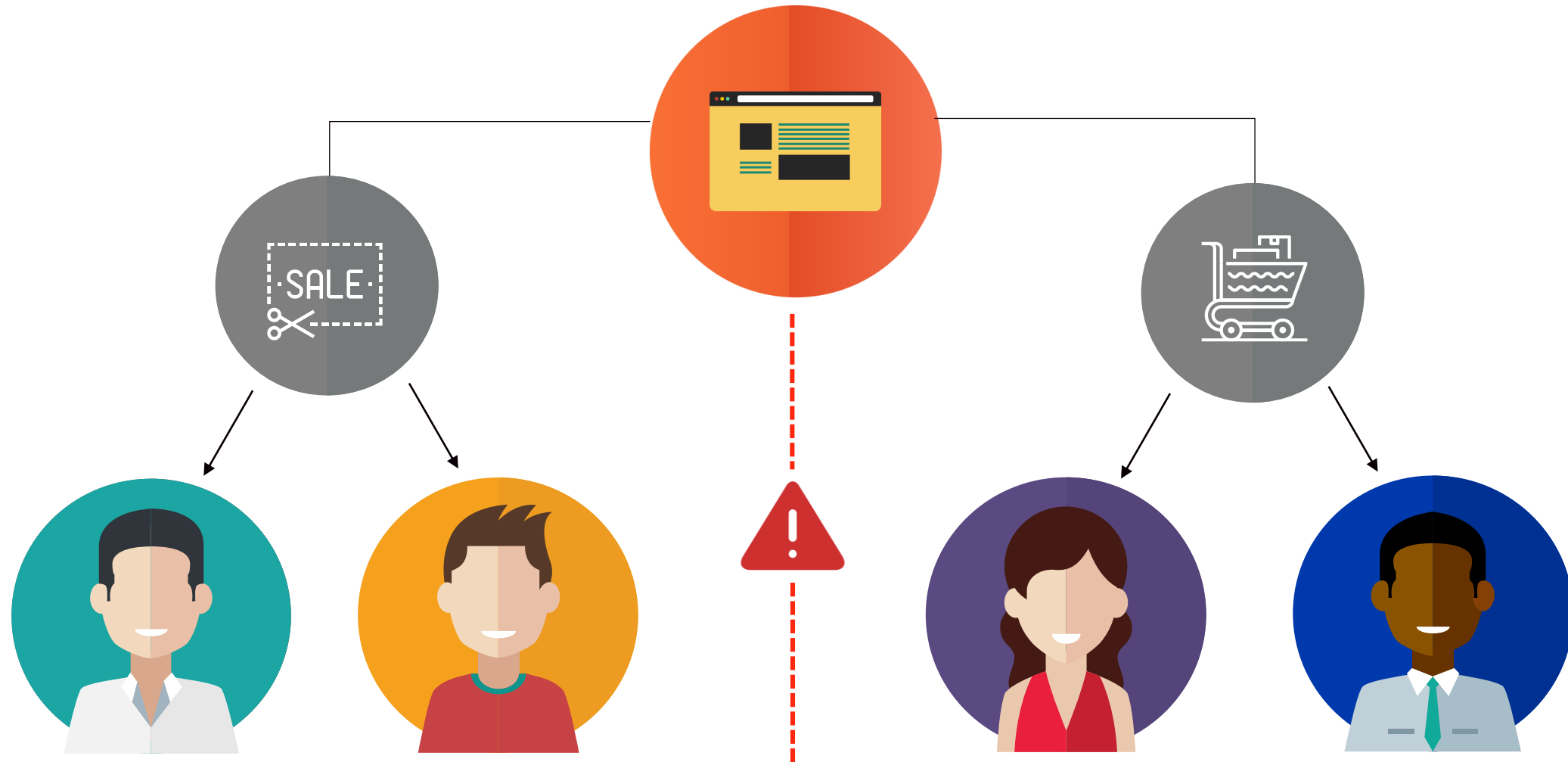


Impact of bias: Exclusion

@StCyrThoughts

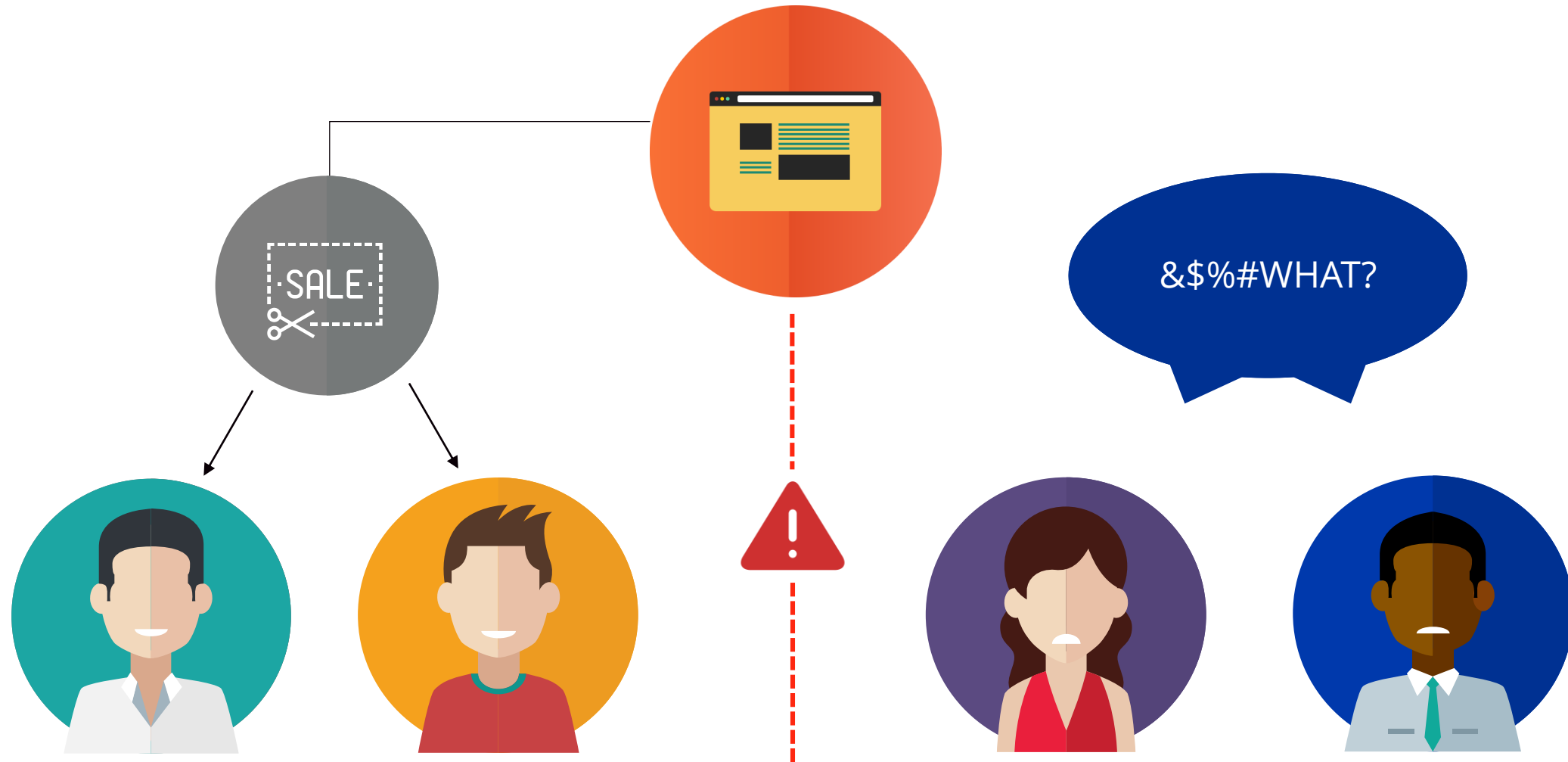




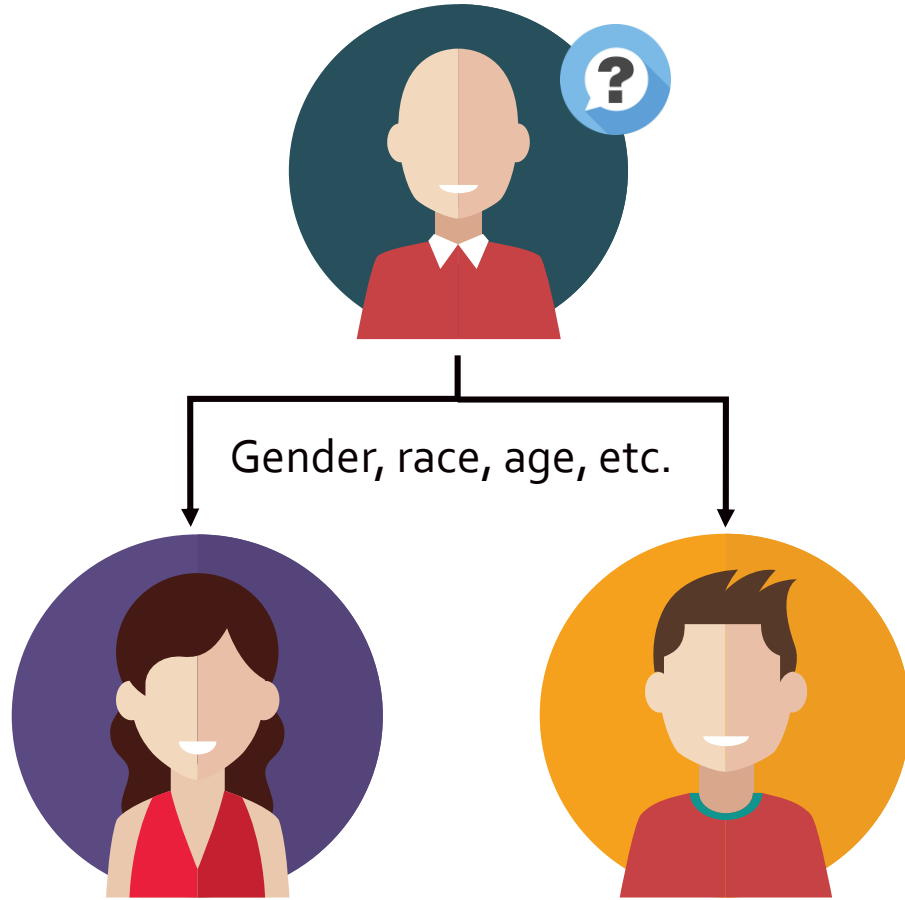


Impact of bias: Exclusion

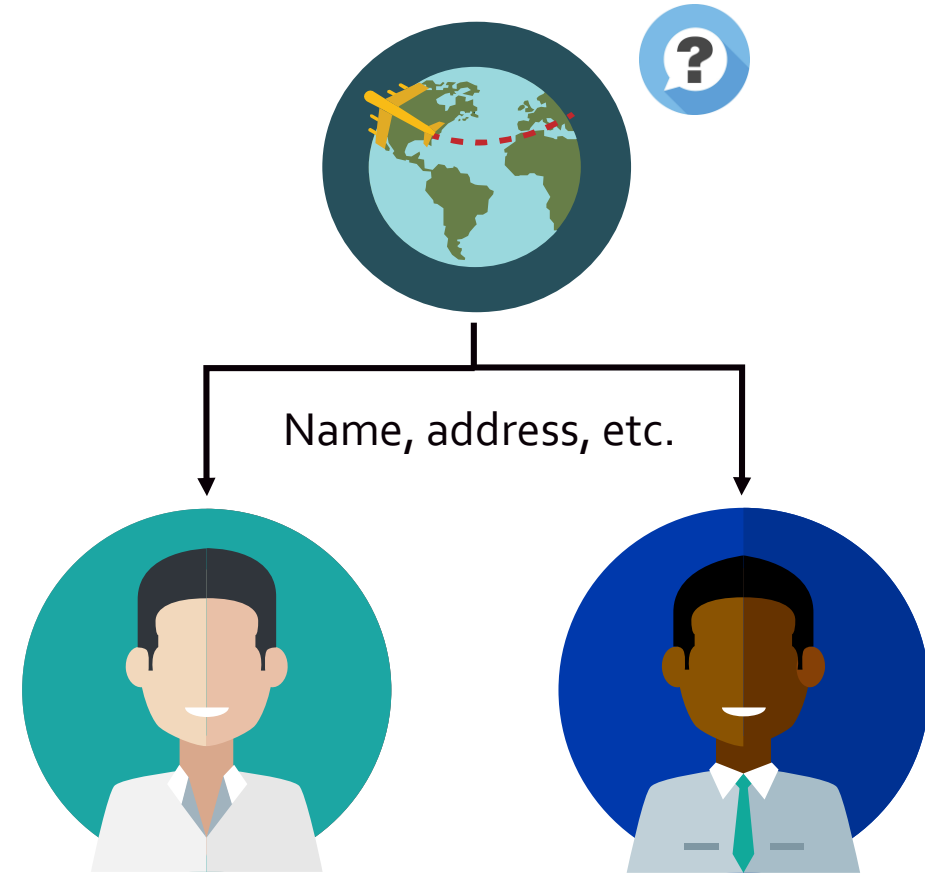
@StCyrThoughts



Direct bias



Embedded bias



**COMPELLING
TALE, CHAP!**



But what can I do about it?

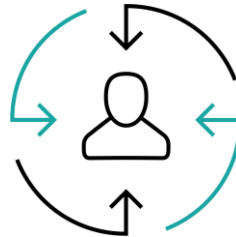
@StCyrThoughts



Integrity

Organization adopts

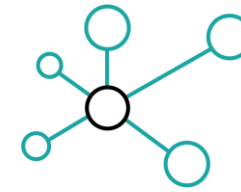
Culture of Integrity



Individualize

Context over

Demographics



Diversity

Data & Team Diversity

Fairness Testing

Integrity

Organization adopts a
Culture of Integrity



Integrity is hard

- Requires shifts in organization culture
- Are we doing the “right thing”?
- Research is your friend!
- Adoption is not your friend 😞



Organizations mimic behavior of their leaders!

Must start from the top down



Has to be OKAY to say things are NOT OKAY

Individuals need to be supported to work in an environment where they can question decisions and actions.



Ultimately YOU are responsible.

But not everyone may agree.



Individualize

Context over
Demographics



Behavior wins.

Model outcomes based on actions, not assumptions based on age, gender, etc.



Track it. Safely.

In order to model after behavior, you need to store behavioral data and profile it.

Make sure you are doing this with privacy in mind!



Blindness is not a cure.

Ignoring demographic data will not fix everything,
but it sure can help!



Context takes time

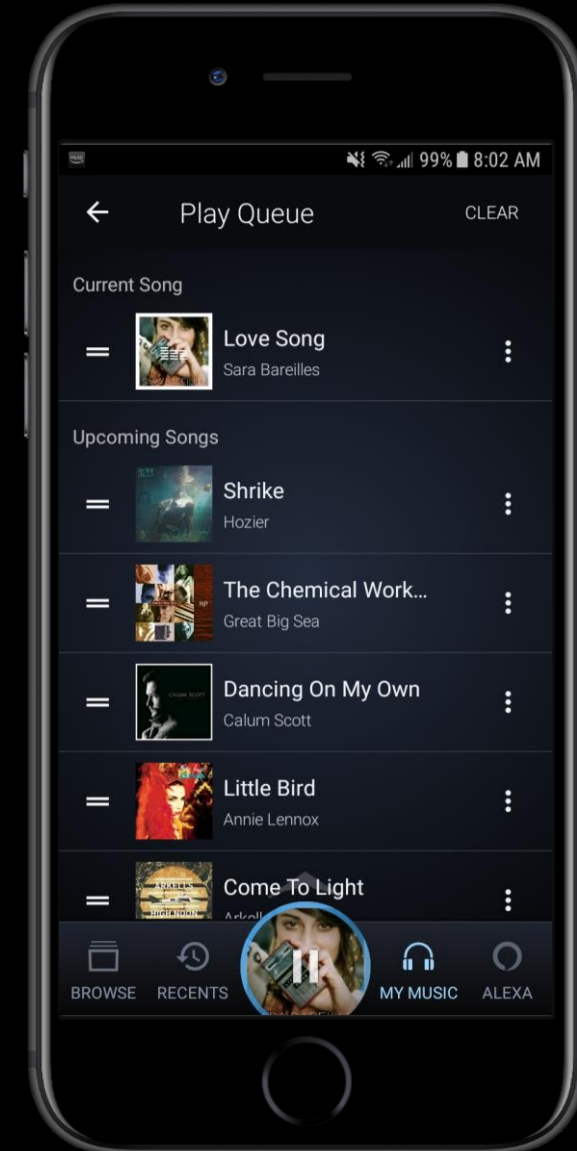
You need to build up behavioral data in order to make predictions.

Existing data sets based on demographics are already out there.



Some fun with demographics!

@StCyrThoughts





Thunderstruck AC/DC



Thunderstruck
AC/DC



Fallin'
Alicia Keys



Choose your playlist

@StCyrThoughts

Thunderstruck AC/DC



Fallin' Alicia Keys



Cleopatra Lumineers



Choose your playlist

@StCyrThoughts

Thunderstruck
AC/DC



Men, 35-50

Fallin'
Alicia Keys



Women, 30-40

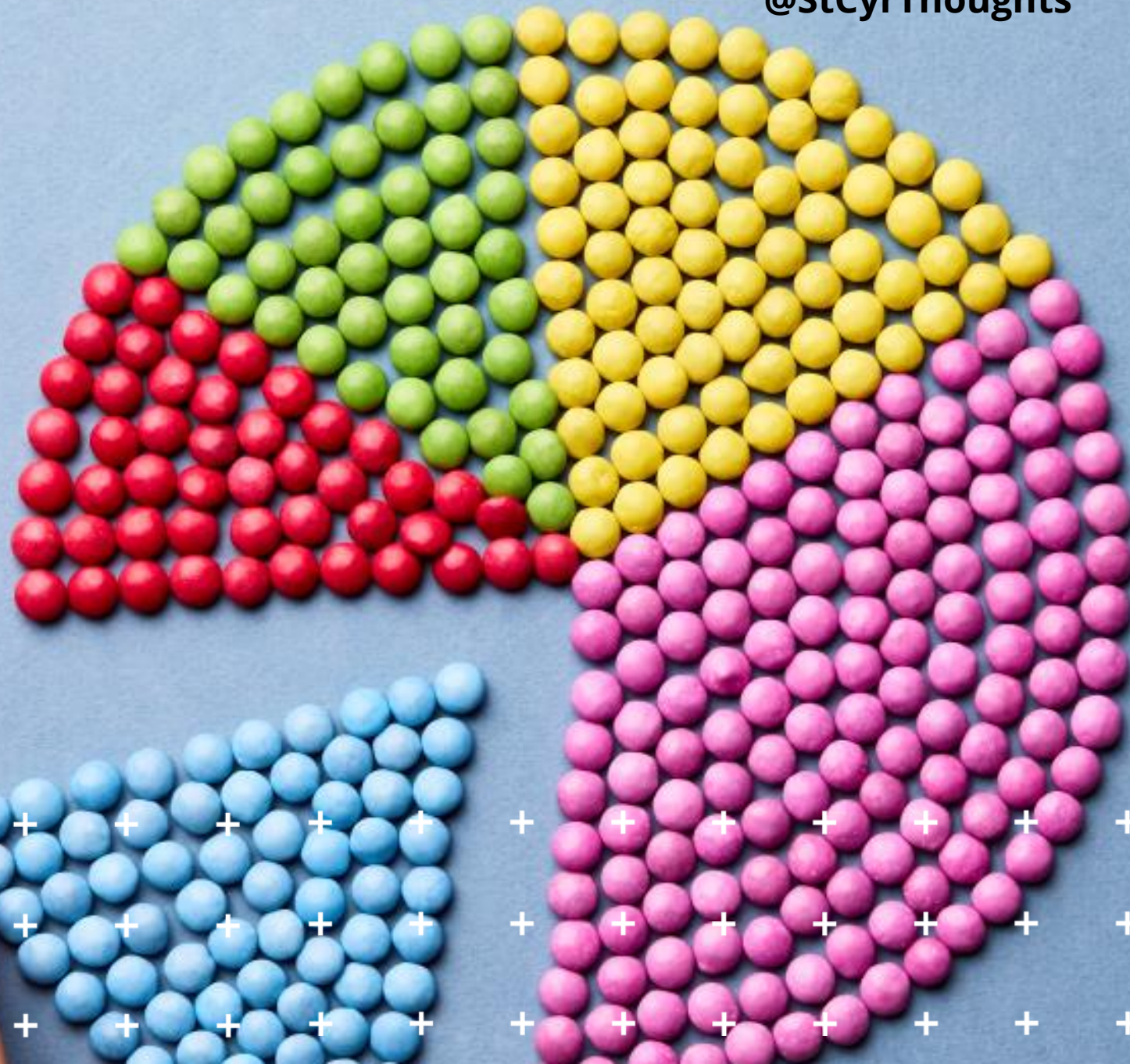
Cleopatra
Lumineers



Hipsters, 25-35

Diversity

Data and Team Diversity
Fairness Testing



Lack of diversity in the team leads to unconscious bias.

Think: Non-binary gender? Physical differences? Cultural differences? Ability differences?



Diverse teams can more easily
break into other markets and
deliver better experiences

Think: China, Japan, Brazil, Africa, etc.



Lack of data diversity leads to invalid predictions

Think: Are *ALL* developers men? Really?



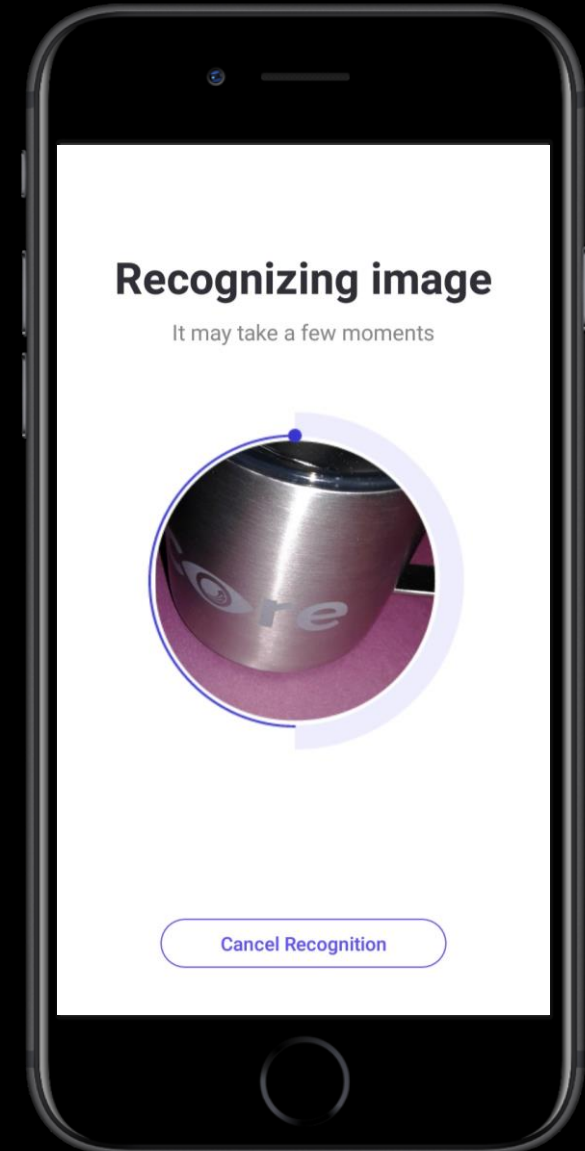
Diverse data allows your algorithm to cover the entirety of your audience and increase accuracy of results

Think: Will they tell the story of when you got it right?



Image Recognition for a 5-Year-Old.

Thank you Angie Jones (@techgirl1908)





Orange



Orange



Apple



Orange



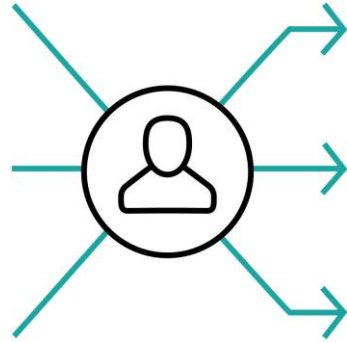
Apple



???

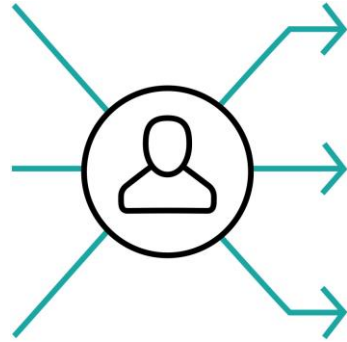
Can you test
for this?





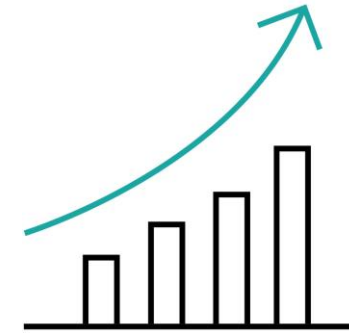
Association Tests

- ✓ Test for unfair, discriminatory, or offensive results caused by input features
- ✓ Looks for statistically significant outcome unfairness
- ✓ Try out sensitive features like Age, Race, Gender



Association Tests

- ✓ Test for unfair, discriminatory, or offensive results caused by input features
- ✓ Looks for statistically significant outcome unfairness
- ✓ Try out sensitive features like Age, Race, Gender



Perturbation Tests

- ✓ Check if small disturbance to feature dramatically changes output
- ✓ Tests if model is highly sensitive to a feature
- ✓ Focus on sensitivity to features like Age, Race, or Gender

ETHICS ALL THE THINGS!



ETHICS ALL THE THINGS!



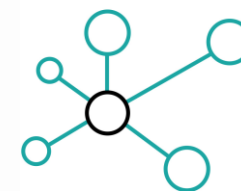
Integrity

Organization adopts
Culture of Integrity



Individualize

Context over
Demographics



Diversity

Data & Team Diversity
Fairness Testing



THANK YOU.

Be excellent to each other!

Jason St-Cyr
@StCyrThoughts

