

# Fair is not the default

[go/fair-not-default](https://www.google.com/go/fair-not-default)

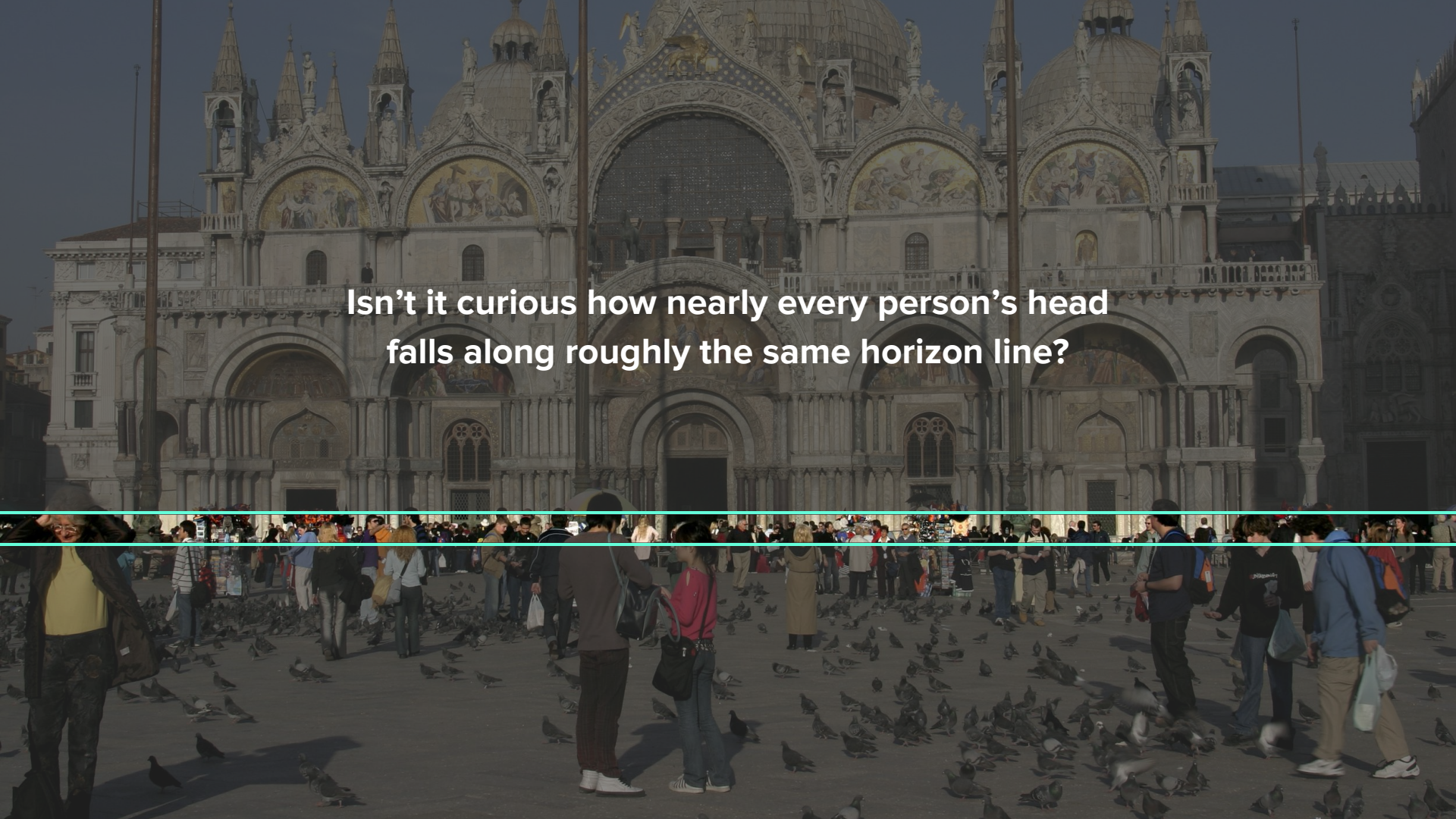
lovejoy@

Google Confidential

A wide-angle photograph of San Marco Square in Venice, Italy. The background is dominated by the ornate facade of St. Mark's Basilica, featuring multiple domes, arches, and intricate mosaics. The square is filled with people, many of whom are interacting with a large number of pigeons on the ground. The scene is captured in bright daylight under a clear blue sky.

# San Marco Square

## Venice

A photograph of St Mark's Basilica in Venice, Italy, taken from a low angle. The cathedral's ornate facade, featuring multiple domes and arches, dominates the background. In the foreground, a large crowd of people is gathered in the Piazza San Marco. The ground is covered with many pigeons. A thin, horizontal cyan line is drawn across the image, passing through the heads of the people in the crowd, illustrating the concept of a common horizon line.

Isn't it curious how nearly every person's head falls along roughly the same horizon line?



It's the result of the photographer being average height. From an *outlier's* perspective, the world might look significantly less uniform.

---

**This is a talk about  
the role of humans in  
machine learning.**

---

---

**But really it's a talk about  
the role of humans in  
decision making.**

---

---

*“It’s true that they can follow instructions at superhuman speed, with superhuman fidelity and over unimaginable quantities of data. **But these instructions don’t come from nowhere.** Although neural networks might be said to write their own programs, they do so towards **goals set by humans, using data collected for human purposes.** If the data is skewed, even by accident, the computers will amplify injustice.”*

— The Guardian

---


---

*“It’s true that they can follow instructions at superhuman speed, with superhuman fidelity and over unimaginable quantities of data. But these instructions don’t come from nowhere. Although neural networks might be said to write their own programs, they do so towards goals set by humans, using data collected for human purposes. If the data is skewed, even by accident, the computers will amplify injustice.”*

— The Guardian

---






—

**Recidivism  
software is  
biased against  
black people**

—

[Machine Bias \(ProPublica\)](#)

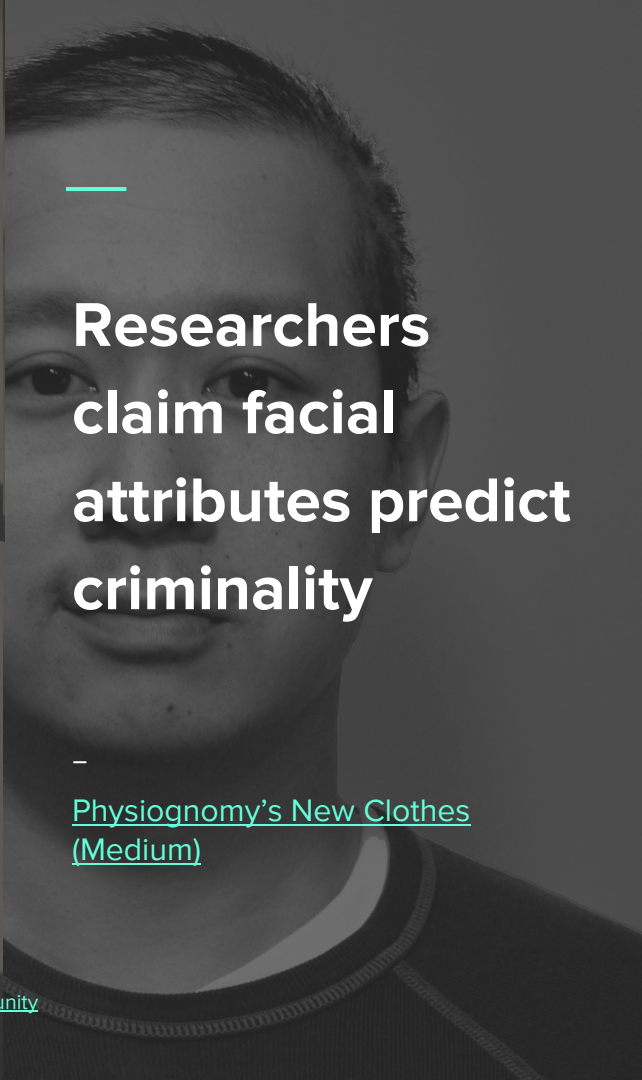


—

**Photo-editing  
app makes faces  
look more  
caucasian**

—

[FaceApp apologizes for building a racist AI \(TechCrunch\)](#)



—

**Researchers  
claim facial  
attributes predict  
criminality**

—

[Physiognomy's New Clothes \(Medium\)](#)


More examples can be found on the [Bias Busting for Machine G+ community](#)

---

When we presume that human judgment can—or should—be removed from the loop, the result is an **unconscious bias network effect.**

And we (Googlers) are just as susceptible to this effect as our users.

---



**Training data are  
collected and  
classified**

CREDIT

[Latent Bias](#), Blaise Agüera y Arcas

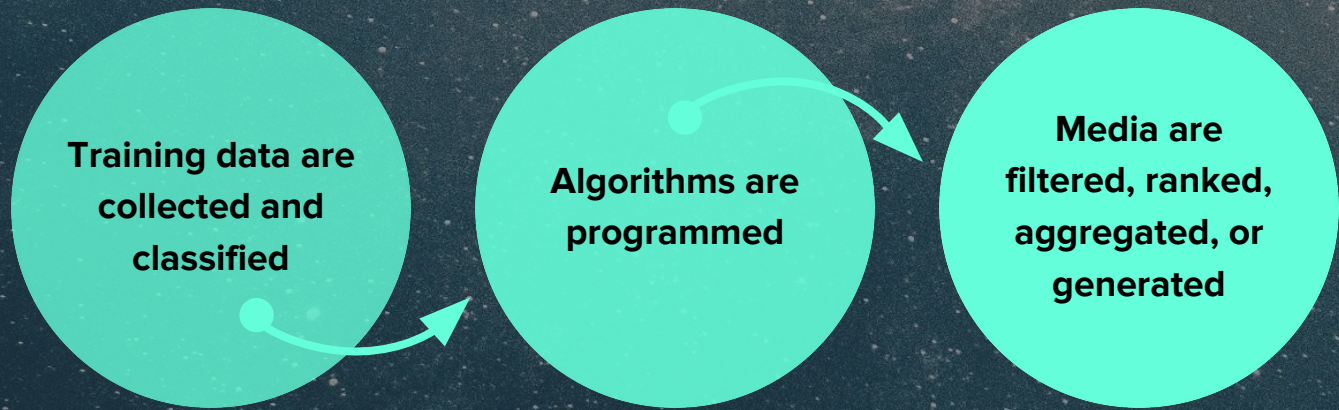
**Training data are  
collected and  
classified**

The diagram consists of two teal circles on a dark blue background with a starry pattern. The left circle contains the text 'Training data are collected and classified'. A teal arrow points from the right side of this circle to the left side of the right circle. The right circle contains the text 'Algorithms are programmed'.

**Algorithms are  
programmed**

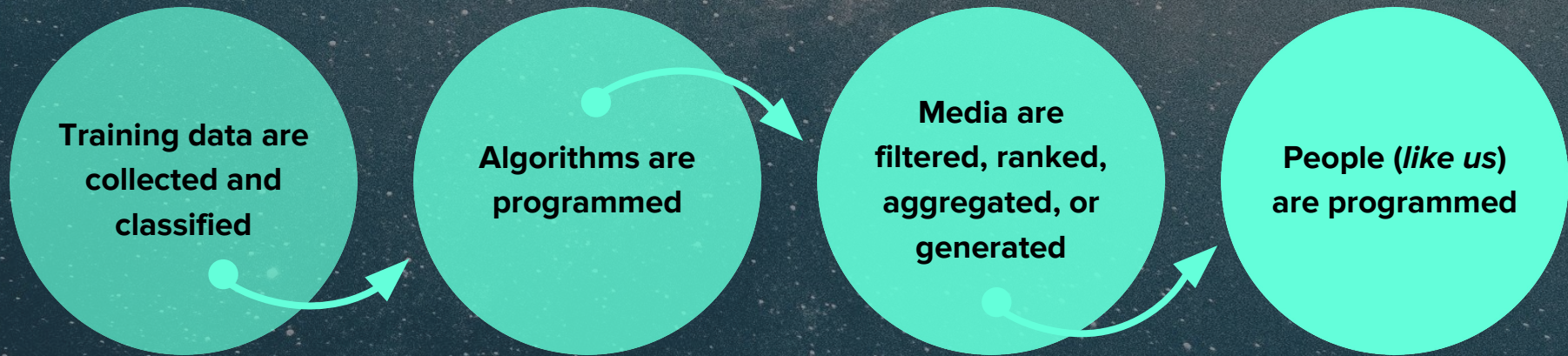
CREDIT

[Latent Bias](#), Blaise Agüera y Arcas



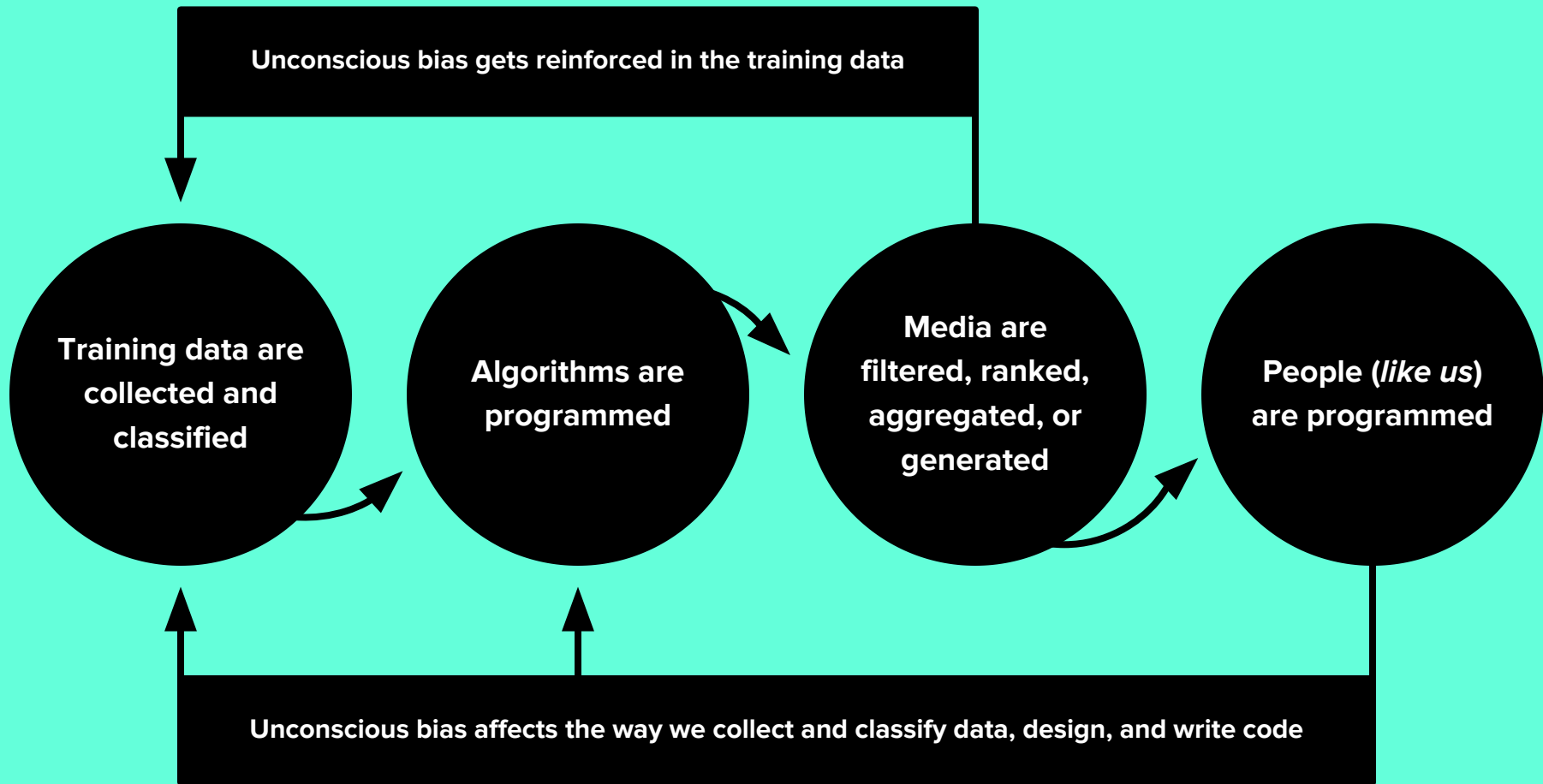
CREDIT

[Latent Bias](#), Blaise Agüera y Arcas



CREDIT

[Latent Bias](#), Blaise Agüera y Arcas



---

If we are going to make the world's information **universally accessible and useful**, we must strive to make the products we are developing—apps, infrastructure, models, and more—**work for everyone**.

---

Learn more at [go/ml-fairness](https://go/ml-fairness)

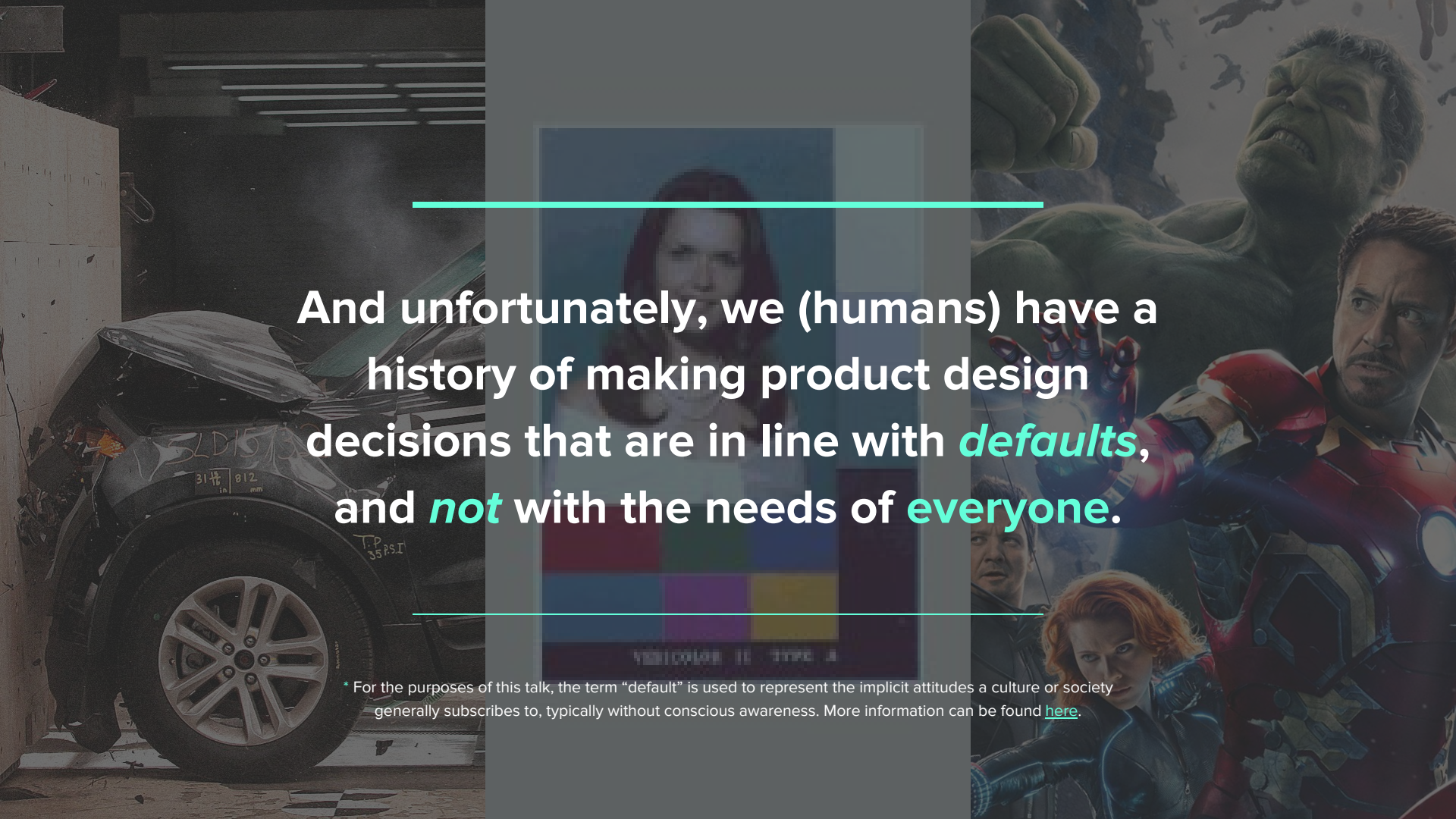


---

If we are going to make the world's information  
universally accessible and useful, we must strive to make  
the products we are developing—apps, infrastructure,  
models, and more—**work for everyone.**

---

Learn more at [go/ml-fairness](https://go/ml-fairness)



And unfortunately, we (humans) have a history of making product design decisions that are in line with *defaults*, and *not* with the needs of *everyone*.

\* For the purposes of this talk, the term “default” is used to represent the implicit attitudes a culture or society generally subscribes to, typically without conscious awareness. More information can be found [here](#).



Female drivers are **47% more likely to be severely injured in an auto accident** because, until 2011, female body-type crash test dummies weren't required by the U.S. Department of Transportation.

SOURCE

[Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk, Bose, Segui-Gomez, and Crandall](#)

[Female dummy makes her mark on male-dominated crash tests \(Washington Post\)](#)



Until the 1990's, virtually all color calibration for color film was based on **caucasian skin tones**. It wasn't until Kodak started receiving complaints from **chocolatiers** and **wood furniture manufacturers** that they began to invest research into supporting a broader range of brown tones.



#### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

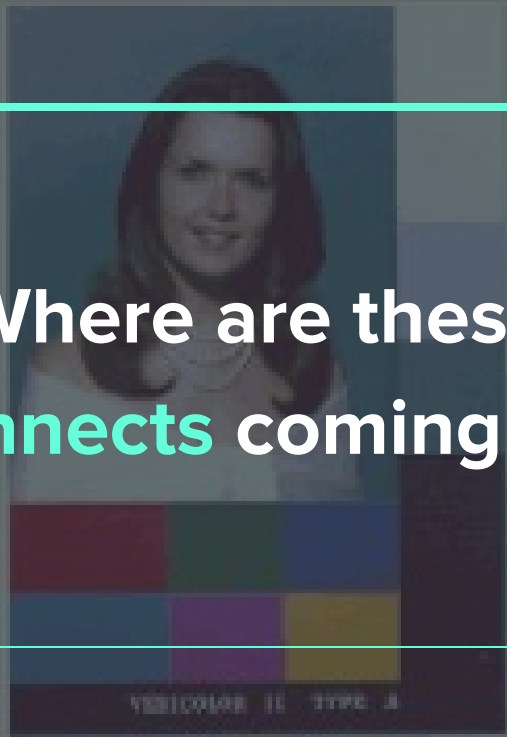
[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)



In the top 100 grossing U.S. live-action films from 2014–2016, male characters were seen and heard nearly twice as often as female characters.

SOURCE

[Geena Davis Institute on Gender in Media](#)

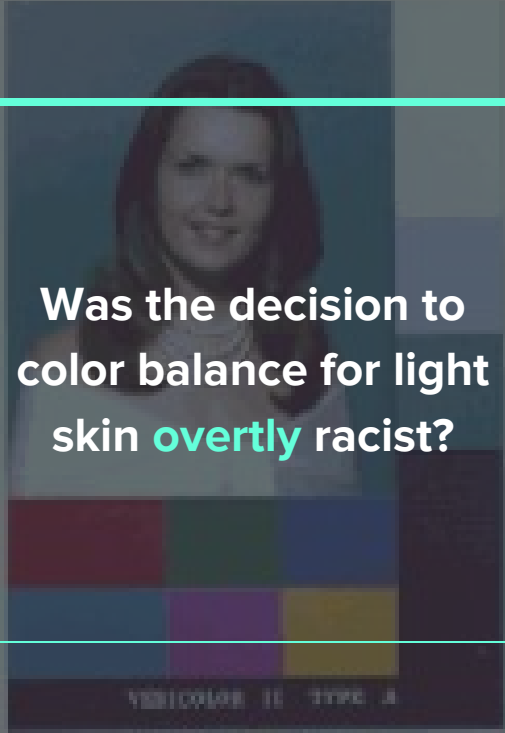


Where are these  
**disconnects** coming from?





Are automakers  
**intentionally** trying to  
harm women?



Was the decision to  
color balance for light  
skin **overtly** racist?



Are creatives in the film  
industry marginalizing  
women **on purpose**?

---

**These are often the behaviors of rational actors making what *seemed* like ‘obvious’ choices, *without* malice or ill-will.**

---

SOURCES

[Why Heuristics Work, Gigerenzer](#)

[Rationality, Blume and Easley](#)

[Recommendations Implicit in Policy Defaults, McKenzie, Liersch, and Finkelstein](#)



---

**Take for example the  
case of “David”**

---





---

**Or was it “Diana”?**

---

The image features a 2x2 grid of baby photos. Each photo shows a baby sitting in a high chair, engaged in a different activity: playing with a stuffed animal, holding a toy, playing with a toy phone, and playing with a toy car. The entire image is overlaid with a dark, semi-transparent grey filter. A horizontal teal line is positioned above the top text, and another is positioned below the bottom text.

**For the purposes of this study, it was both.**

Participants watched a series of videos of babies.  
For each video, **half of the participants were told  
a boy's name, the other half a girl's name.**

SOURCES

[Sex Differences: A Study of the Eye of the Beholder, Condry and Condry](#)  
[Steven Pinker & Elizabeth Spelke | The Science of Gender & Science](#)

---

**When the babies did something unambiguous, reports were not affected by the perceived gender.**

If the baby clearly smiled, for example, everyone said the baby was smiling or happy.

---

SOURCES

[Sex Differences: A Study of the Eye of the Beholder, Condry and Condry](#)  
[Steven Pinker & Elizabeth Spelke | The Science of Gender & Science](#)



---

Then the babies played with a jack-in-the-box toy. When it suddenly popped up, the child was startled and jumped backward.

---



SOURCES

[Sex Differences: A Study of the Eye of the Beholder, Condry and Condry](#)  
[Steven Pinker & Elizabeth Spelke | The Science of Gender & Science](#)

---

**“She's afraid”**

---

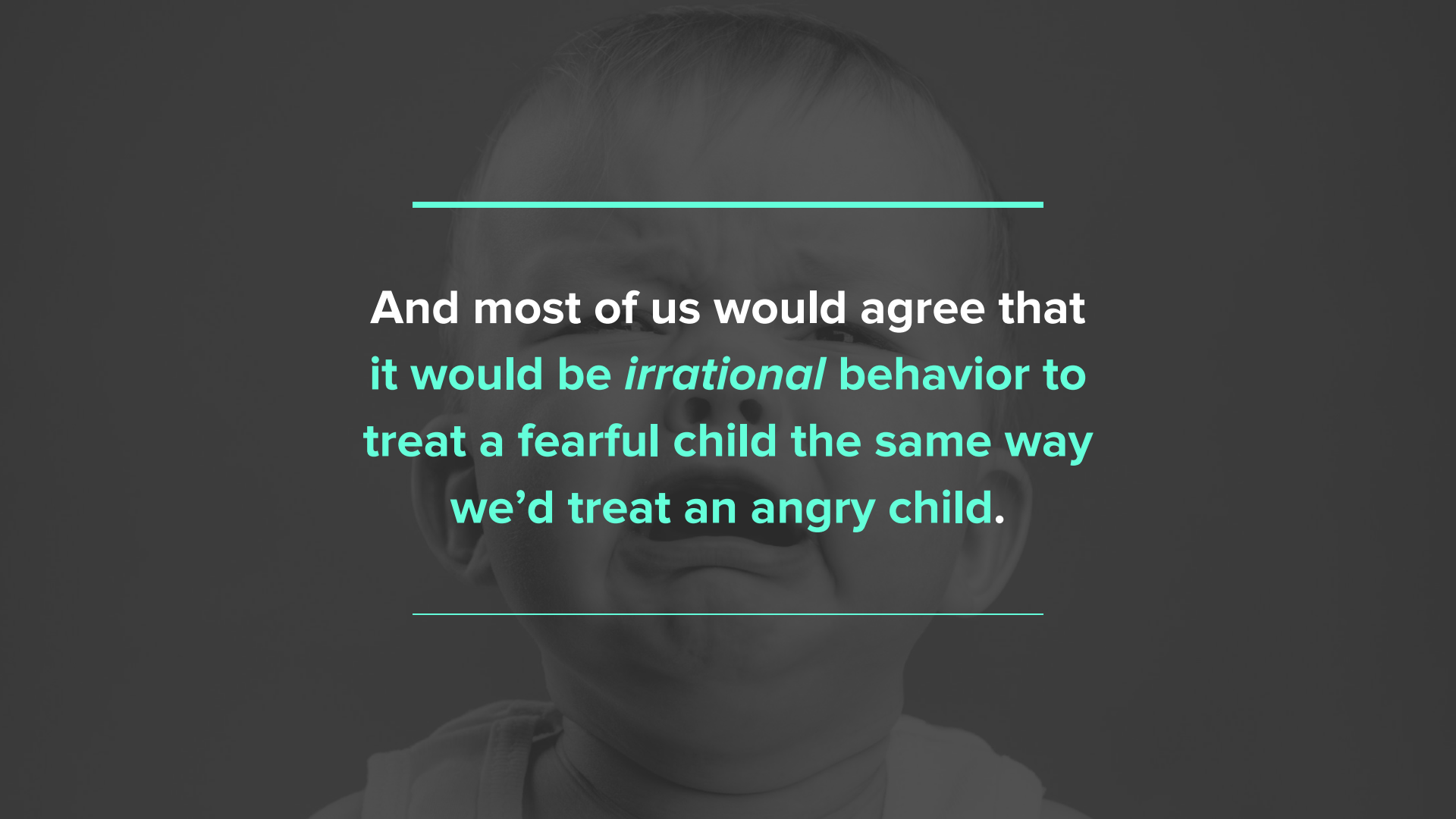
---

**“He's angry”**

---

**SOURCES**

[Sex Differences: A Study of the Eye of the Beholder, Condry and Condry](#)  
[Steven Pinker & Elizabeth Spelke | The Science of Gender & Science](#)



---

And most of us would agree that  
it would be *irrational* behavior to  
treat a fearful child the same way  
we'd treat an angry child.

---

---

*“If knowledge of a child's gender affects adults' perception of that child, then male and female children are going to elicit different reactions from the world, different patterns of encouragement. **These perceptions matter**, even in parents who are committed to treating sons and daughters alike.”*

— Elizabeth Spelke

---

SOURCES

[Steven Pinker & Elizabeth Spelke | The Science of Gender & Science](#)



---

# Same child, same reaction, different perception.

---

## REPLICATION AND RELATED STUDIES

[The Gender Stereotyping of Emotions: Plant, Hyde, Keltner, Devine](#)

[A comparison of observed and reported adult-infant interactions: Effects of perceived sex, Culp, Cook, Housley](#)

[Adult perceptions of the infant as a function of gender labeling and observer gender, Delk, Madden, Livingston, Ryan](#)

[Surprising Smiles and Unanticipated Frowns: How Emotion and Status Influence Gender Categorization, Smith, LaFrance, Knol, Moes](#)

---

**Human perception** drives  
virtually every facet of  
machine learning.

---

---

**I propose we\* make machine-learning  
intentionally human-centered and  
intervene for fairness.**

---

\* We = Humans. This isn't something any one company should be doing in isolation, but we're in a good position to start.

---

# Tenets

Designing for fairness

---

---

01

# Be Accountable

**We can't take our hands  
off the steering wheel.**

In rejecting the myth of neutral data,  
we are committing to be more  
conscious and conscientious.

---

01

# Be Accountable

Present day

↳ Robots

↳ Yada yada yada

↳ Future

Present day

↳ **Humans**

↳ **Still humans**

---

02

# Be Skeptical

**Challenge assumptions at every turn.**

We can't blindly rely on the systems that underpin conventional wisdom.

---

02

# Be Skeptical

## Journalism

[Fake news is indistinguishable](#)

## Customer reviews

[Male reviews skew average scores](#)

## Standardized tests

[SAT scores don't predict grades](#)

## Medical science

[Experiments over-recruit whites](#)

## Crime statistics

[Racial profiling is real](#)



---

03

# Be Humane

**Success metrics should  
bring out the best in  
human nature.**

Standard engagement metrics confine people to whatever they've done *before*, rather than empowering what they're capable of doing *next*.

---

03

# Be Humane

**Time well spent**

[timewellspent.io](http://timewellspent.io)

**Learning and expression**

**Exploration and connection**

**User-defined goals**

---

04

# Be Humble

**We don't *always* know better.**

The tech industry is in love with “disruption”, but frequently that means imposing a vision of the future *onto* users and expecting them to adapt.

---

04

# Be Humble

*“There are two ways to get people used to automation. The soft and fuzzy way, ... just keep reassuring people until they’re comfortable. And then there’s the second way: let the humans **take control when they need to.**”*

— [NPR](#)

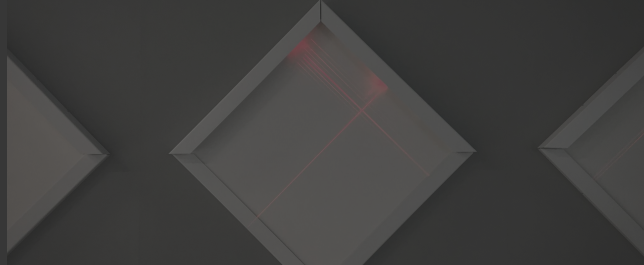
---

**Augmenting** → NOT → **Automating**

**Supporting** → NOT → **Replacing**

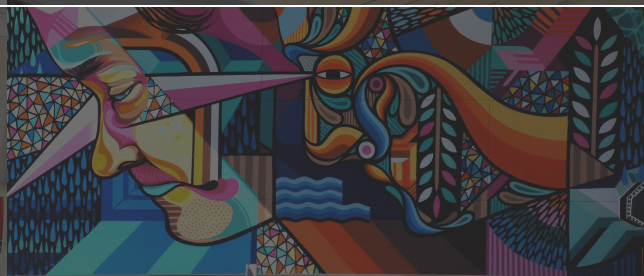
---

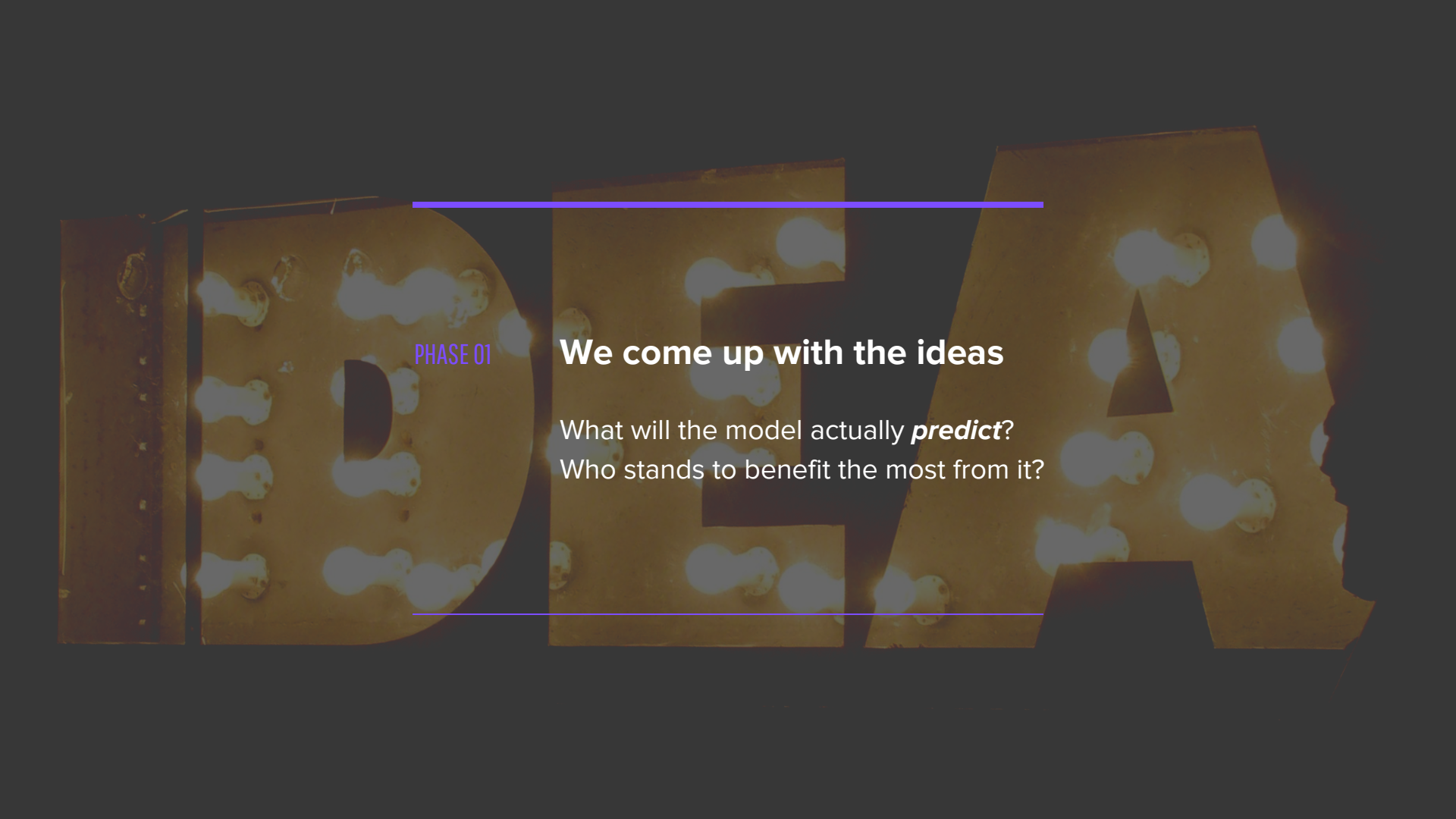
# IDEA



# In practice

An ML user journey in 12 phases





PHASE 01

## We come up with the ideas

What will the model actually *predict*?  
Who stands to benefit the most from it?

PHASE 02

## We define the ML strategy

Why would heuristics be less effective than a machine-learned model?





---

PHASE 03

## We conduct the market research

How do users describe the benefits of ML?  
What reference points are they bringing?

---

PHASE 04

## We design the experiences

How are people solving these problems today?  
How might ML improve things?

---

PHASE 05

## We collect the data

What factors determine if the data are representative?

---

PHASE 06

## We design the Rater protocols

Hypothetically speaking, would this be an unambiguous task for end-users to perform?

*(Author's note: While the use of Raters is primarily for supervised learning, labeled data—ground truth or otherwise—have a wide variety of applications, so I chose to include this as a prominent phase. Reinforcement learning is likely the only place they're entirely absent.)*

PHASE 06

## Sidebar on Raters

**Speed and Agreement** are the bedrock measures of “[click-workers](#)”.

The general idea is that if a lot of people can perform many quick tasks, the sheer volume of consensus will balance their lack of individual expertise. But without careful consideration for the diversity of Raters, click-work turns into exponential groupthink; baking cultural biases directly into training data.

---

PHASE 07

## We train the Raters

How will you verify that Raters are performing tasks 'correctly'?

---

V

VI

---

PHASE 08

## We try to label all the things

Where did *you* acquire the knowledge necessary to successfully label these data?

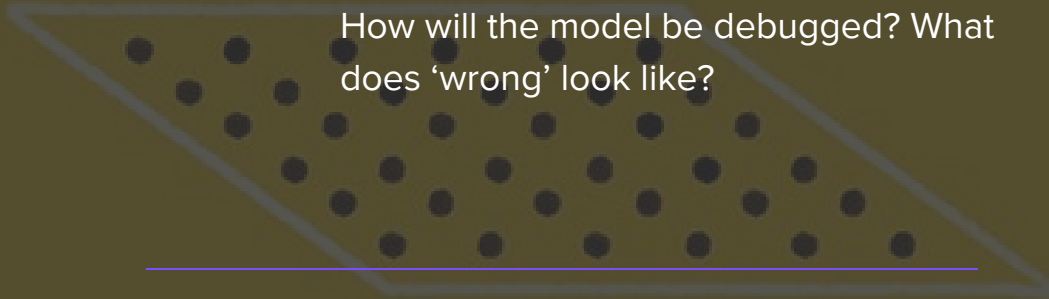
---



PHASE 09

## We train the models

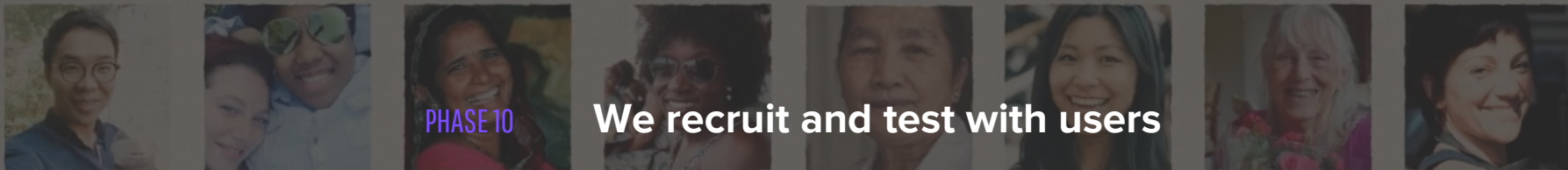
How will the model be debugged? What does 'wrong' look like?



ACTIVATED  
NEURONS

INPUT  
LAYER

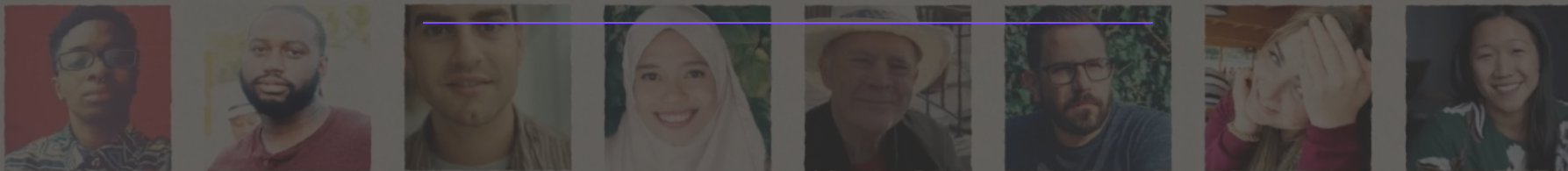
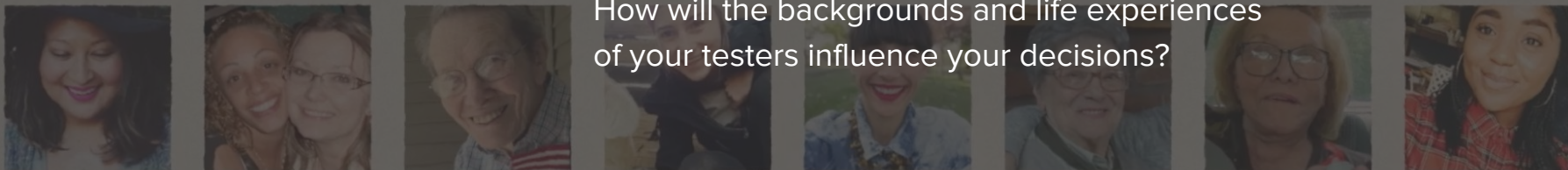


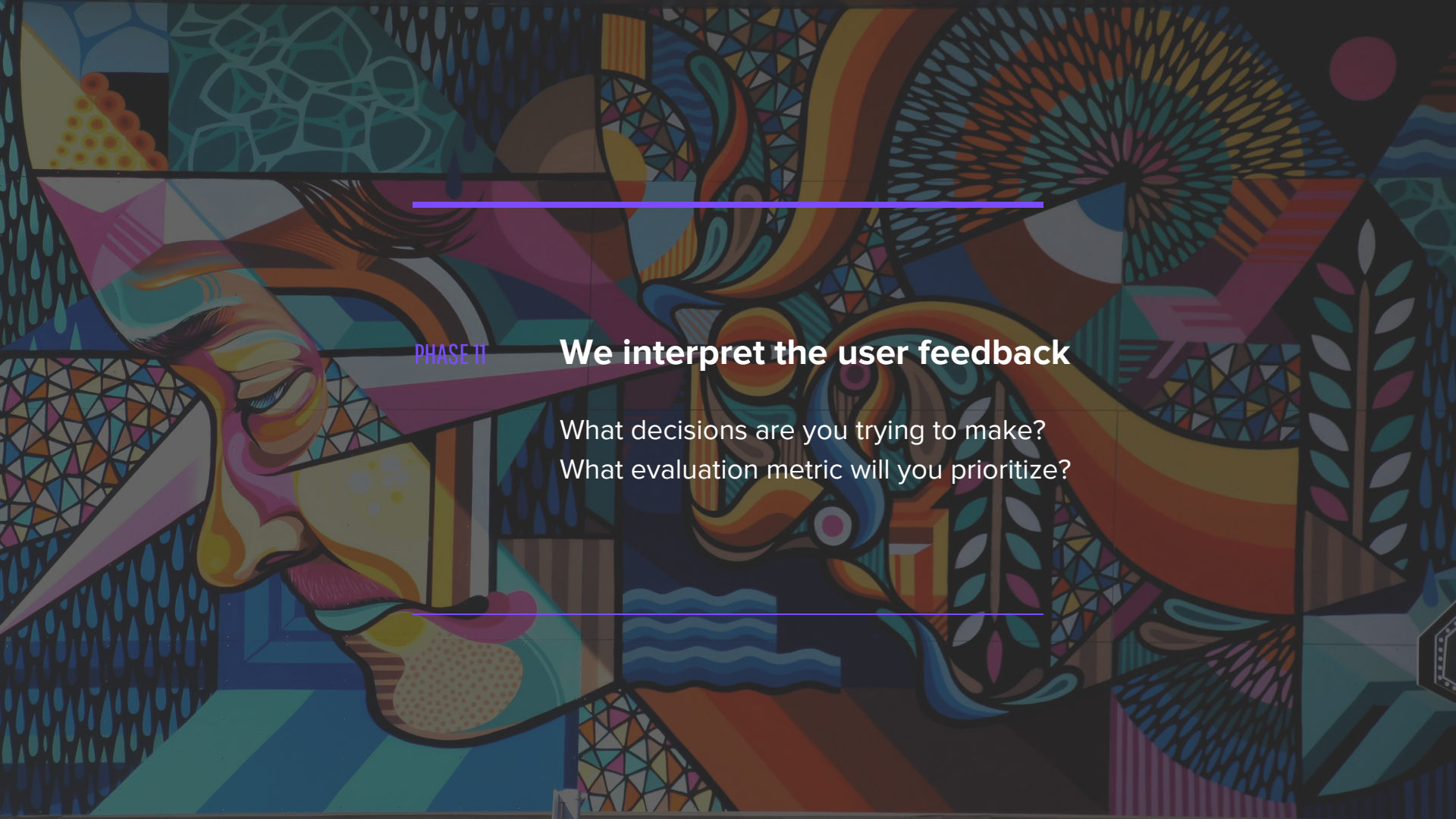


PHASE 10

## We recruit and test with users

How will the backgrounds and life experiences of your testers influence your decisions?





PHASE II

## We interpret the user feedback

What decisions are you trying to make?

What evaluation metric will you prioritize?

PHASE 12

## We craft the PR

How will users see themselves reflected  
in marketing materials and demos?



---

Finally, I'd like to offer **3 human-centered diagnostics** when designing with ML.

If you're finding it tricky to answer these, it might be a signal to slow down and take a closer look.

---

---

01

## If a human were to perform this task, what would ‘appropriate’ social behavior look like?

What interpersonal cues might be relevant that are missing from your input or interface?  
E.g. body language, tone of voice.

Take **autocomplete** for example: In what context would it be acceptable to finish another person’s sentence before they've stopped talking?

Bear in mind that no one culture is capable of representing universal norms, especially for social interactions, so we need to be mindful of our intuitions. [The psychological effects of algorithmic discrimination likely mirror those of social discrimination.](#)

autocomplete is an|



autocomplete is an **interruption**

---

02

## If a human were to perform this task, might we call it an expression of their personality?

Tasks that have unambiguous utility, are perceived as repetitive or boring by users, and/or benefit from super fast response times are ideal candidates for ML augmentation.



Reply to all

Great idea!

I like that idea.

Please like me.

### **Grey area: Suggesting a reply in an email or SMS**

likely has a priming effect; impacting the user's response even if they don't use it. And the fact that suggestions are even offered may lead the recipient to question the authenticity of the sentiment.

**Augmentation:** The goals of a **self-driving car** are unambiguous, and the benefits of a computer's superhuman reaction time offer objective utility. No one (hopefully) would say a driver got into an accident because they were expressing themselves.

---

# 03

## Who are you?

... OK, now what do your data teach you about **everyone else?**

Our traits don't necessarily define us, but it's foolish to pretend we don't see them. By taking the potentially uncomfortable step of inventorying these traits—physical, social, cognitive, and otherwise—we're **getting proximate to those who are reminded of their differentness every time a 'default' is invoked in day-to-day life.**

### I am...

- |                                      |                       |
|--------------------------------------|-----------------------|
| White                                | Male                  |
| A parent                             | Affluent              |
| Visually impaired                    | Physically capable    |
| Agnostic                             | Culturally Jewish     |
| Insured                              | A homeowner           |
| Urban-dwelling                       | Not college educated  |
| In my 30's                           | Married               |
| Cisgender                            | Heterosexual          |
| 99% percentile height                | In good mental health |
| A speaker of "standard" U.S. english |                       |

---

03

Because hopefully that  
can help your world look  
a bit **less like this...**



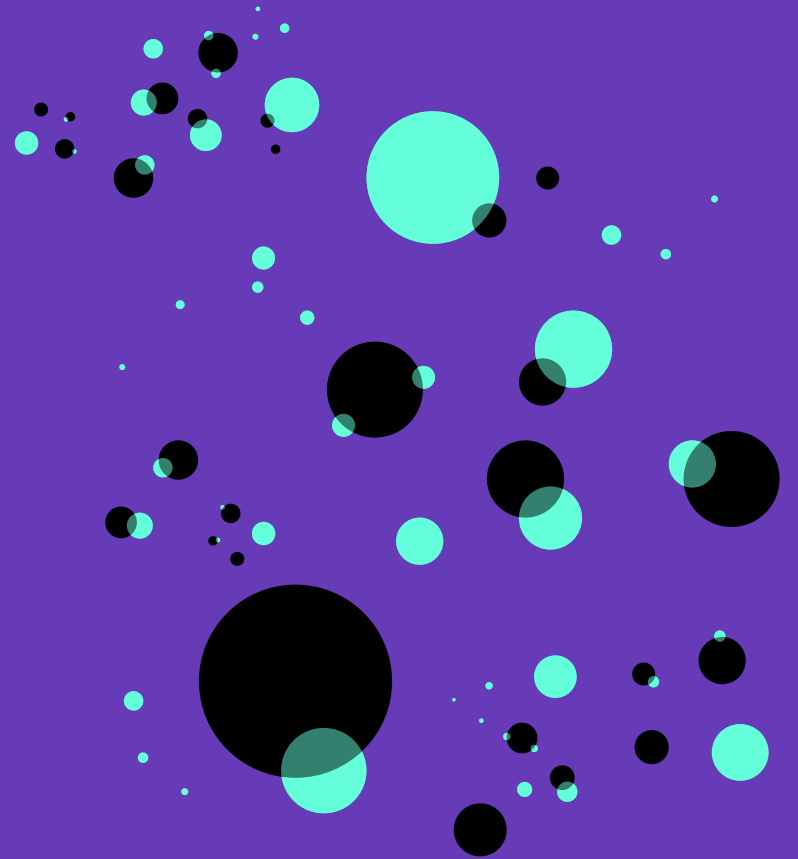


---

03

Because hopefully that  
can help your world look  
a bit **less like this...**

...and a lot **more like this**



---

# Thank you

Dive deeper at [go/ml-fairness](https://go/ml-fairness)

---

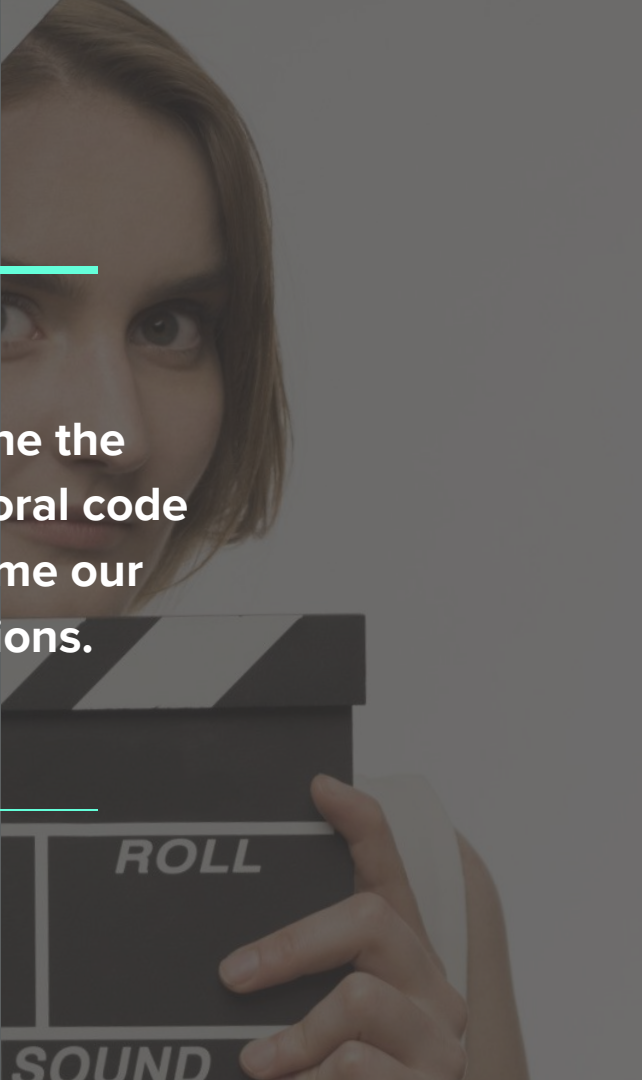
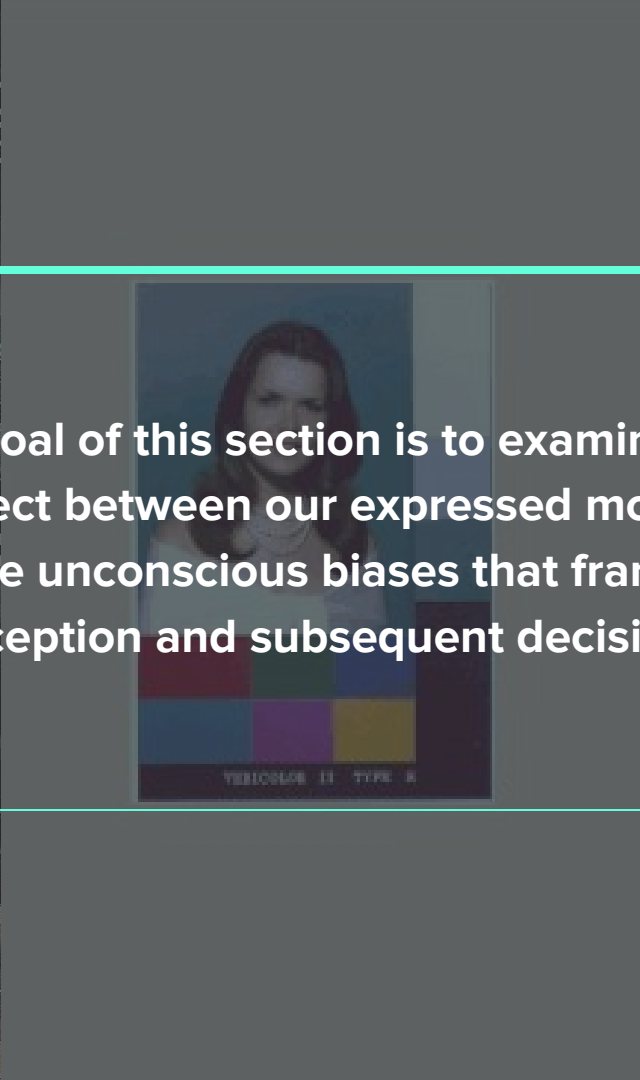
---

# Appendix

---



The goal of this section is to examine the disconnect between our expressed moral code and the unconscious biases that frame our perception and subsequent decisions.



A woman with light brown hair is holding a black clapperboard with white text and diagonal stripes. The clapperboard has fields for 'SCENE', 'TAKE', 'ROLL', 'DATE', and 'SOUND'. A semi-transparent dark grey overlay covers the image, with a thin cyan horizontal line above and below the text. The text '01 Gender roles in film' is centered on the overlay.

01 **Gender roles  
in film**

## GENDER ROLES IN FILM

---

In 2015, women shared top billing with men in the top four grossing live-action films in the U.S.

**How much were female characters seen and heard compared to men in those films?**

---



## GENDER ROLES IN FILM

---

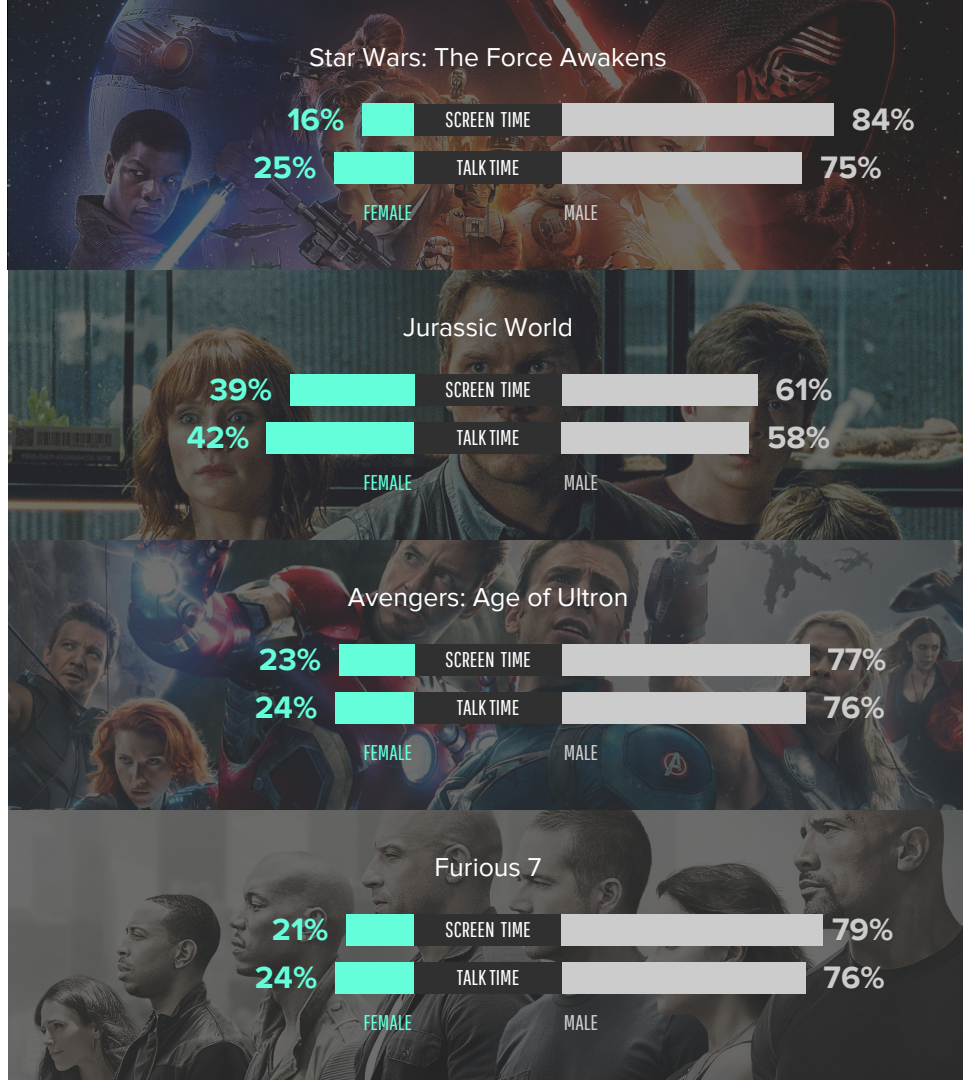
In 2015, women shared top billing with men in the top four grossing live-action films in the U.S.

**How much were female characters seen and heard compared to men in those films?**

---

SOURCE

[Geena Davis Institute on Gender in Media](#)



## GENDER ROLES IN FILM

---

**The numbers improved slightly for the top grossing films featuring female leads.**

But still contain some surprisingly disproportionate numbers.

---





## GENDER ROLES IN FILM

---

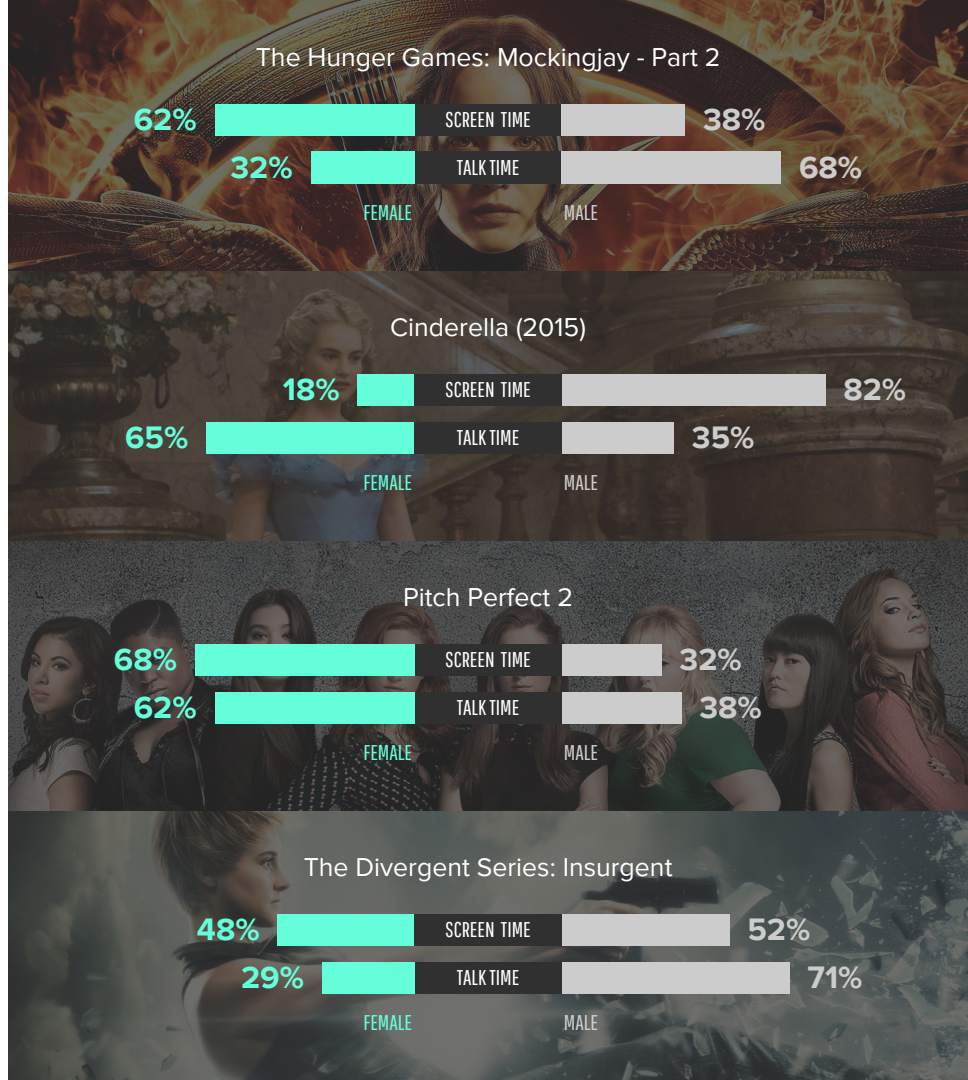
The numbers improved slightly for the top grossing films featuring female leads.

But still contain some surprisingly disproportionate numbers.

---

SOURCE

[Geena Davis Institute on Gender in Media](#)

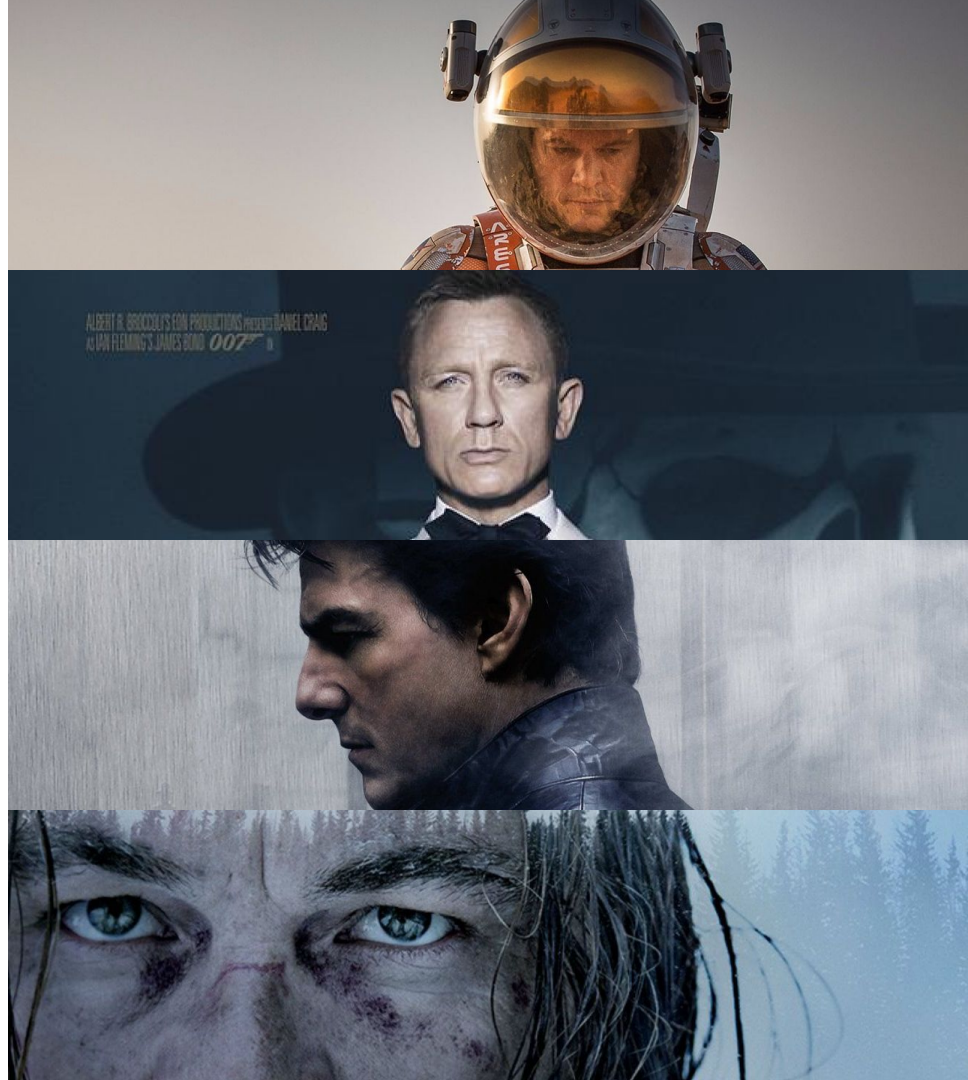


## GENDER ROLES IN FILM

---

**But the gap increased substantially for the top grossing films featuring male leads.**

---



## GENDER ROLES IN FILM

---

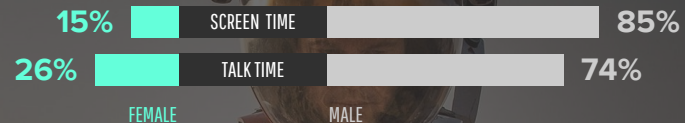
But the gap increased substantially for the top grossing films featuring male leads.

---

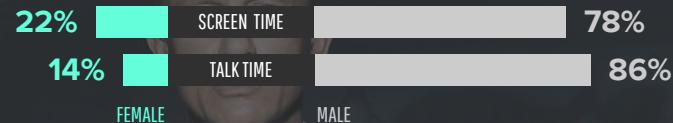
SOURCE

[Geena Davis Institute on Gender in Media](#)

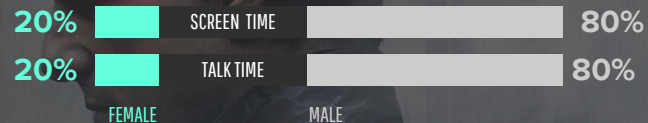
### The Martian



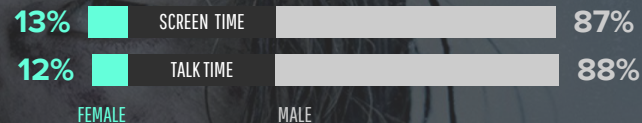
### Spectre




### Mission: Impossible - Rogue Nation



### The Revenant



A promotional image for the movie Avengers: Age of Ultron. The main cast members are shown in their superhero costumes: Iron Man, Thor, Hulk, Captain America, Black Widow, and Hawkeye. They are standing in a line against a dark, smoky background with many small, dark silhouettes of flying creatures. The lighting is dramatic, highlighting the characters' faces and costumes.

GENDER ROLES IN FILM

---

Overall, for the top 100 grossing  
live-action films from 2015 in the U.S.

Male characters were **seen**

Male characters were **heard**

**1.84x**

**1.78x**

▲ more than female characters

▲ more than female characters

---

SOURCE

[Geena Davis Institute on Gender in Media](#)

## Related research

When women and men speak the exact same amount, women are perceived to be speaking

**22%**  
more than men

When in a mixed-gender group, women speak less than men until they comprise

**80%**  
of the group

In making the top 250 grossing films of 2015 in the U.S., women held

**19%**  
of behind-the-scenes creative roles

---

### SOURCES

[Speaker sex and perceived apportionment of talk, Cutler and Scott](#)

[Gender Inequality in Deliberative Participation, Karpowitz Mendelberg and Shaker](#)

[Behind-the-Scenes Employment of Women on the Top 100, 250, and 500 Films of 2015, Lauzen](#)

---

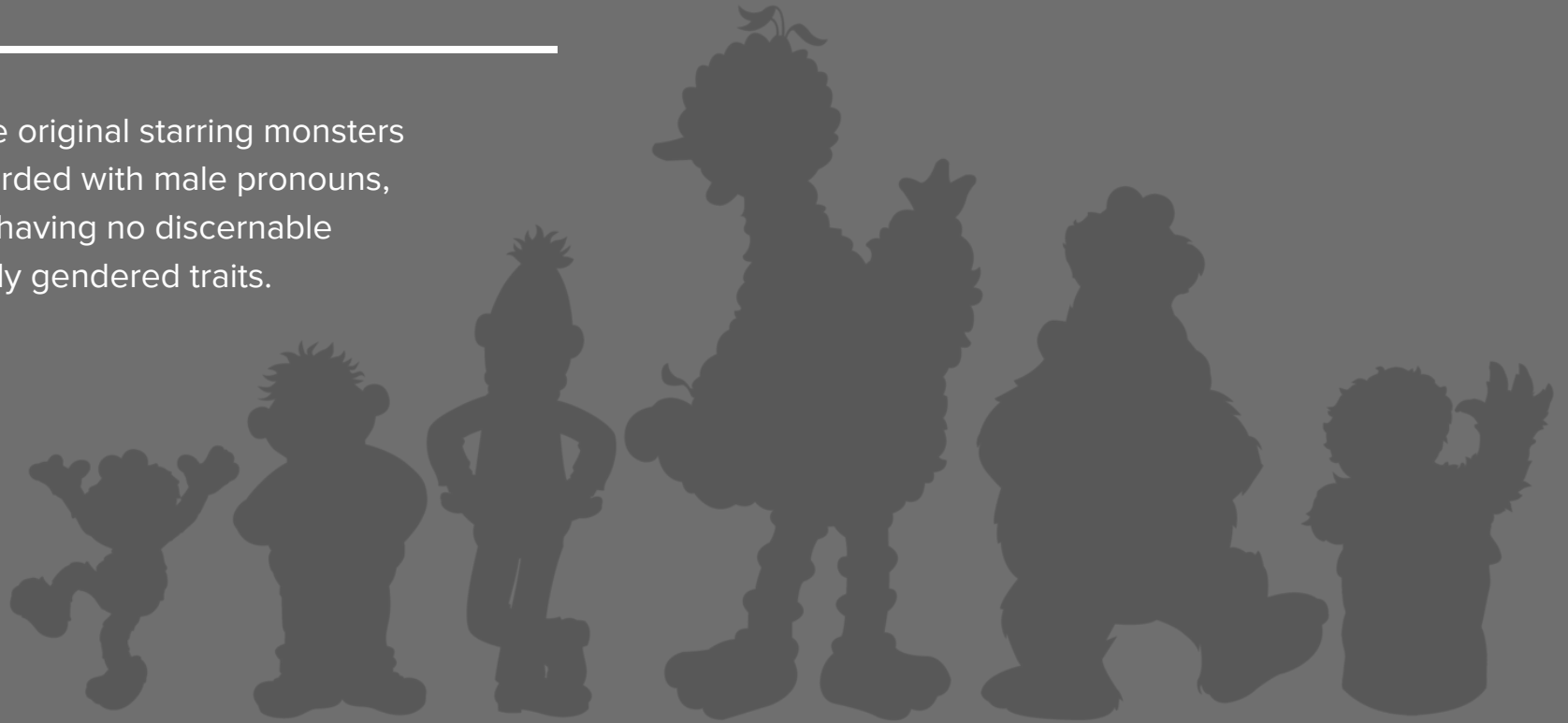
## One more observation

Have you ever noticed the “gender”  
makeup of Sesame Street?



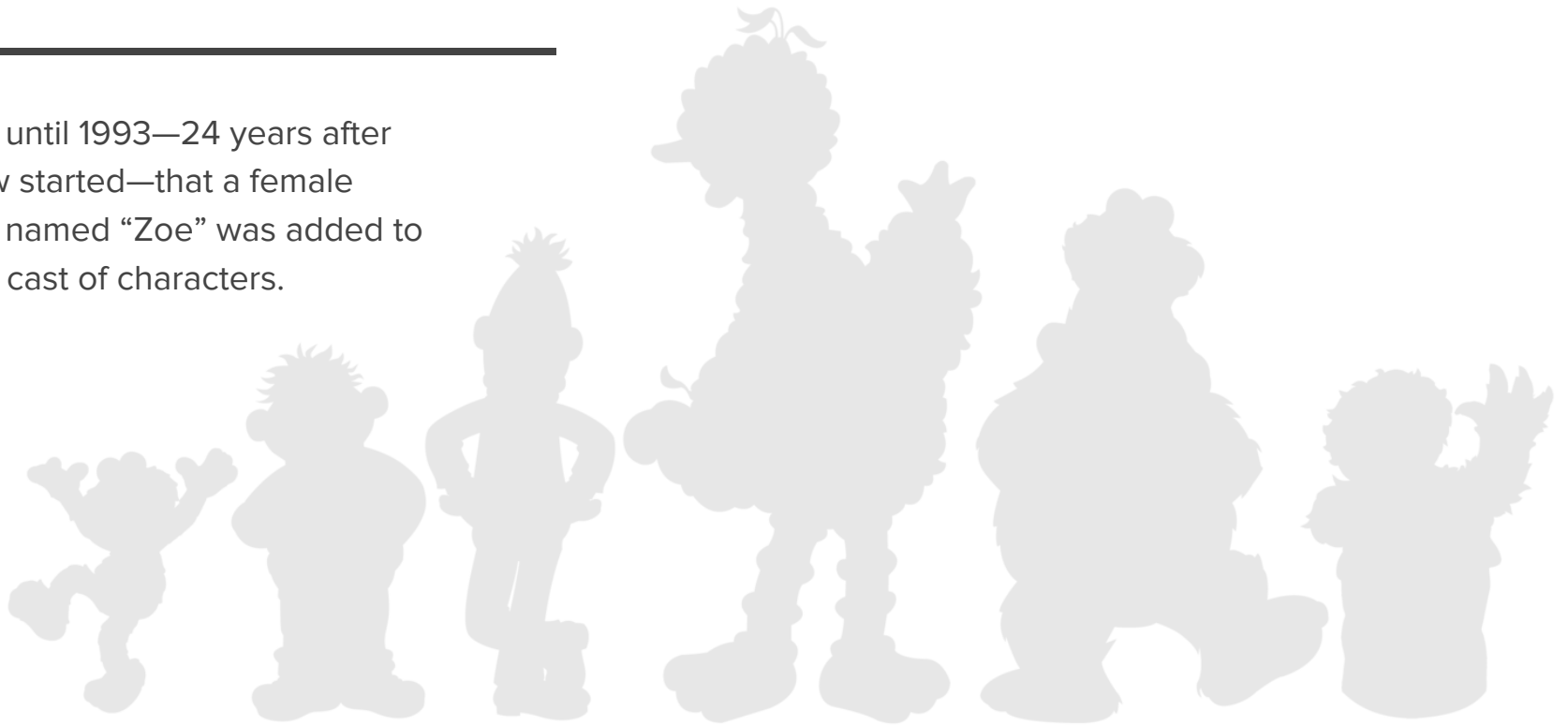
---

All of the original starring monsters are regarded with male pronouns, despite having no discernable physically gendered traits.



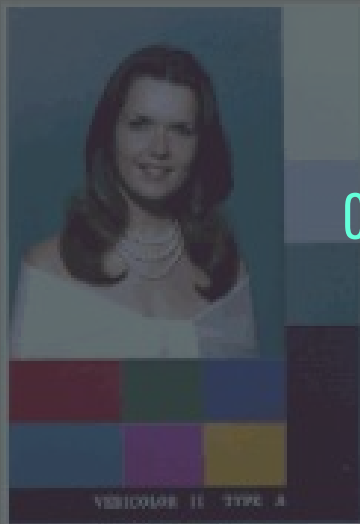
---

It wasn't until 1993—24 years after the show started—that a female monster named “Zoe” was added to the core cast of characters.





## 02 Skin tone in photography



Kodak

## SKIN TONE IN PHOTOGRAPHY

---

### This is a “Shirley Card”

Named after a Kodak studio model named Shirley Page, they were the primary method for calibrating color when processing film.

---

#### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[How Kodak's Shirley Cards Set Photography's Skin-Tone Standard, NPR](#)



## SKIN TONE IN PHOTOGRAPHY

---

**Until about 1990, virtually all Shirley Cards featured caucasian women.**

---

### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)



## SKIN TONE IN PHOTOGRAPHY

---

**As a result, photos featuring people with light skin looked fairly accurate.**

---

### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)



Film  
Year  
Credit

Kodachrome  
1970

[Darren Davis, Flickr](#)

## SKIN TONE IN PHOTOGRAPHY

---

Photos featuring people with darker skin, not so much...

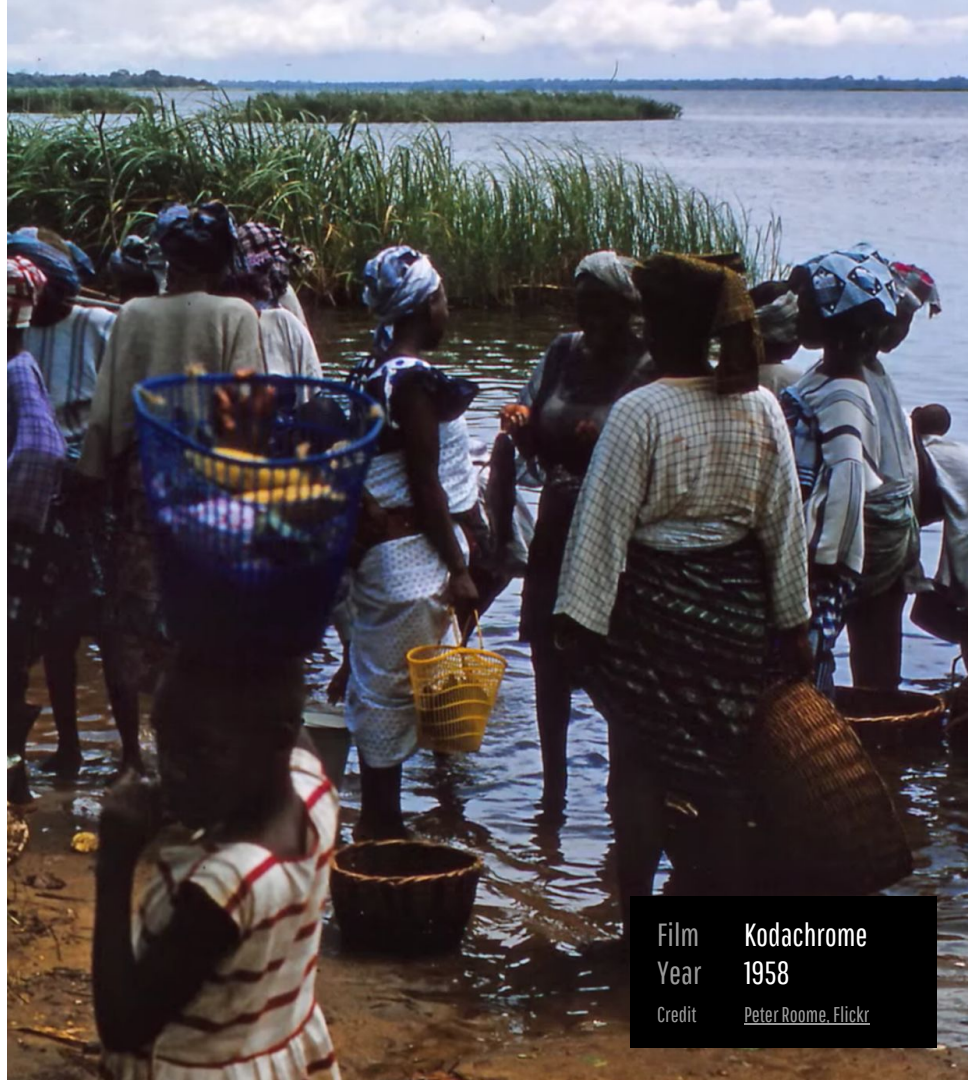
---

### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)



Film  
Year

Kodachrome  
1958

Credit

[Peter Roome, Flickr](#)

## SKIN TONE IN PHOTOGRAPHY

---

**And when there was a mix, the difference was most noticeable.**

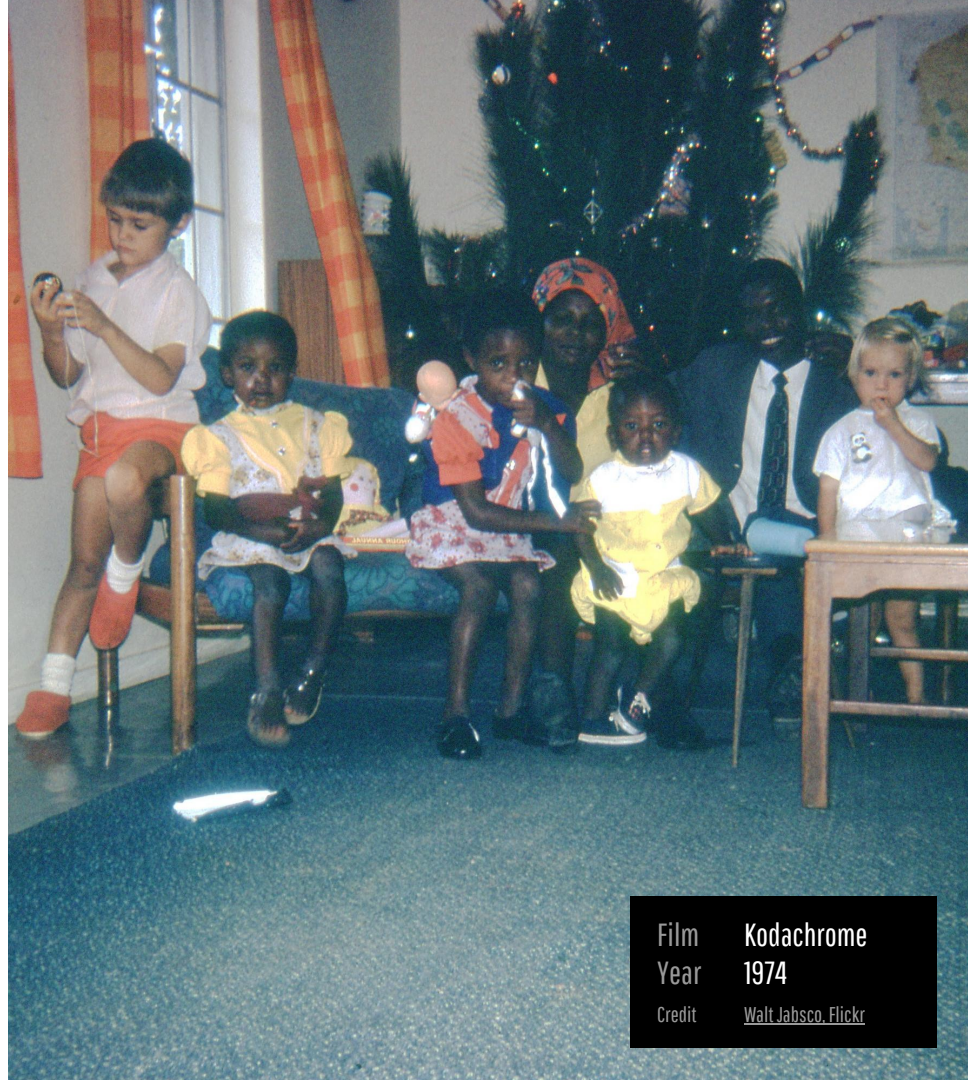
---

### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)



Film  
Year

Kodachrome  
1974

Credit

[Walt Jabsco, Flickr](#)

## SKIN TONE IN PHOTOGRAPHY

---

**As society became more integrated, photographers found workarounds.**

The most common techniques were to shoot with significantly brighter lights (making the room really hot!), using a stronger flash (42% brighter!), and preparing separate cameras with different calibrations for people with different skin tones.

---

### SOURCES

[Colour Balance, Image Technologies, and Cognitive Equity](#), Roth  
[How Photography Was Optimized for White Skin Color](#) (Priceonomics)



Title	In the Heat of the Night
Year	1967



Title	The Whoopi Goldberg Show
Year	1992-1993

## SKIN TONE IN PHOTOGRAPHY

---

**But motivation for innovation came from chocolatiers and wood furniture manufacturers.**

Kodak was receiving complaints that they weren't getting the right brown tones on chocolates, and that stains and wood grains were not true to life.

---

### SOURCES

[Color film was built for white people. Here's what it did to dark skin. \(Vox\)](#)

[Colour Balance, Image Technologies, and Cognitive Equity, Roth](#)

[How Photography Was Optimized for White Skin Color \(Priceonomics\)](#)





SKIN TONE IN PHOTOGRAPHY

## Related research

As of 2014, across all 2 million Implicit Association Test (IAT) participants

**51%**

showed a moderate to strong bias for white faces

As of 2014, across all U.S. white IAT participants, the median "D score" was

**0.402**

indicating a moderate bias for white faces

As of 2011, of adults working in Engineering roles in the U.S. (age 16 and over)

**4.8%**

were black

SOURCES

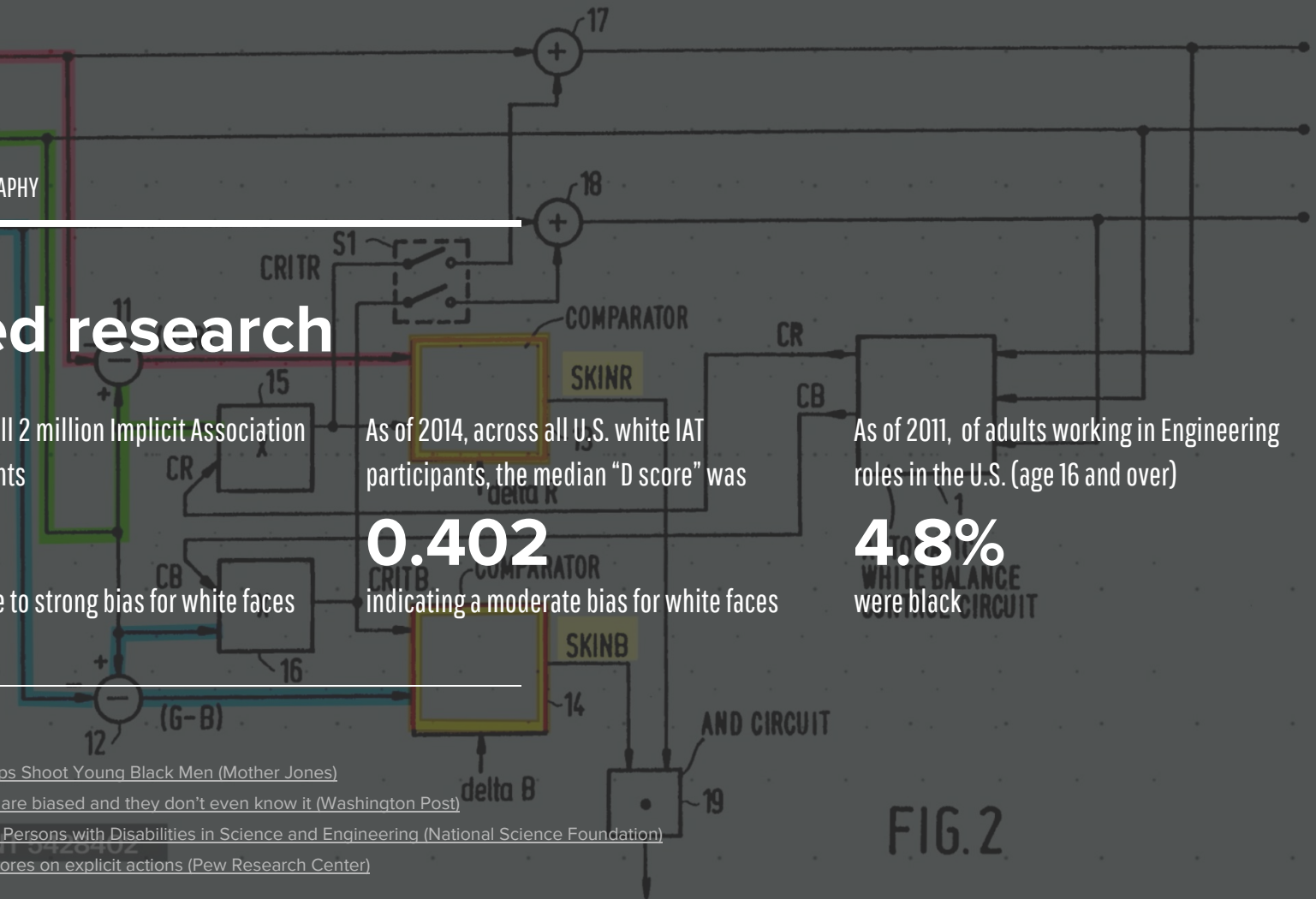
[The Science of Why Cops Shoot Young Black Men \(Mother Jones\)](#)

[Across America, whites are biased and they don't even know it \(Washington Post\)](#)

[Women, Minorities, and Persons with Disabilities in Science and Engineering \(National Science Foundation\)](#)

APPENDIX: [Effects of IAT scores on explicit actions \(Pew Research Center\)](#)

FIG. 2





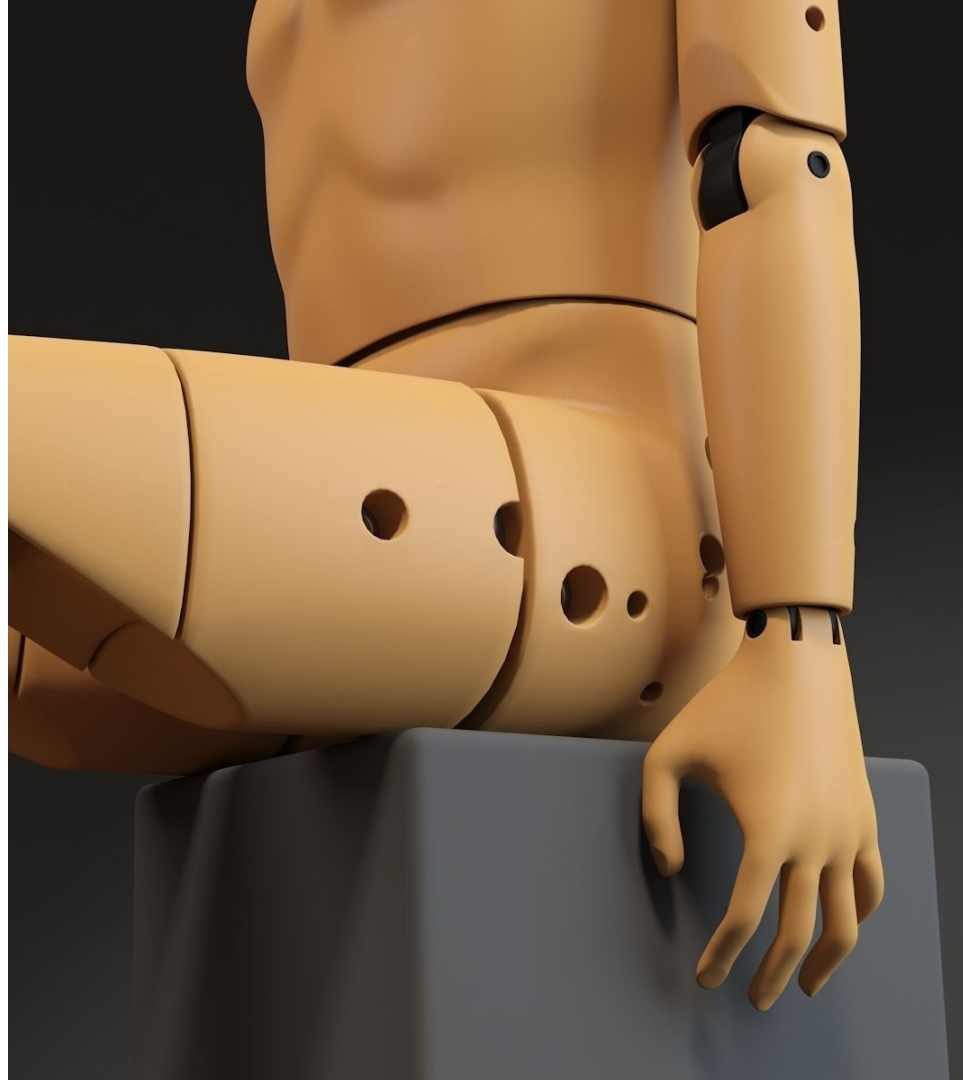
03 **Automotive  
safety**

**Until 2011, female body-type crash test dummies were not required by the United States Department of Transportation.**

---

SOURCE

[Female dummy makes her mark on male-dominated crash tests \(Washington Post\)](#)



## AUTOMOTIVE SAFETY

---

**As a result, female drivers are at a higher risk behind the wheel.**

Odds a female driver will sustain severe injuries in an accident

**47%**

▲ higher than a male driver

---

### SOURCE

[Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk, Bose, Sequi-Gomez, and Crandall](#)



Things are improving, but the target percentile for female test dummies remains problematic.

Male body percentile

**50th**

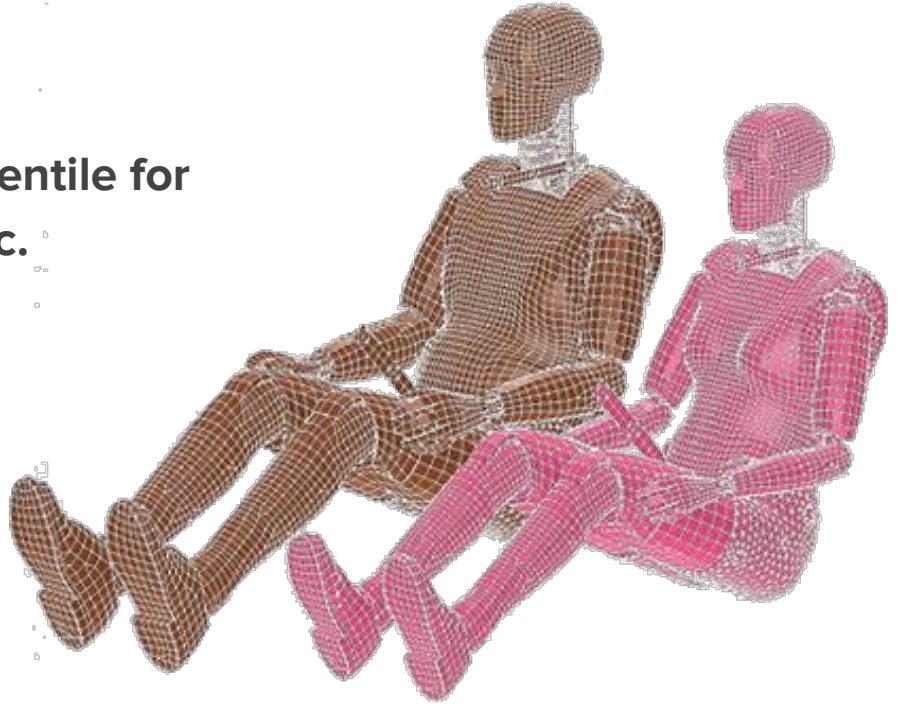
5'9" 176 lbs

Female body percentile

**5th**

4'11" 108 lbs

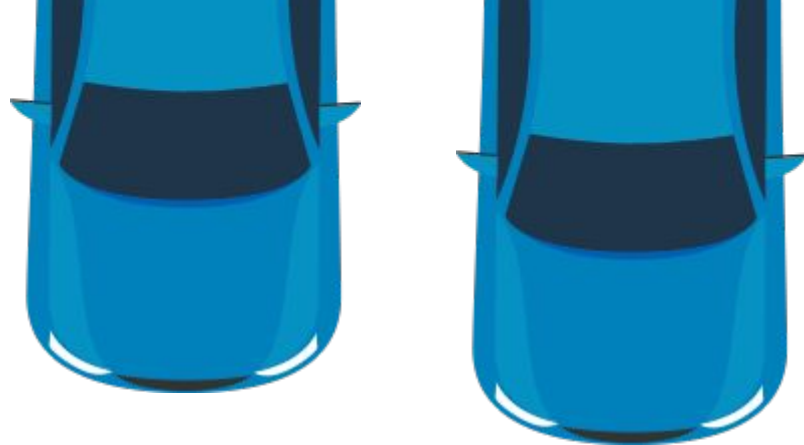
---



SOURCE

[Female dummy makes her mark on male-dominated crash tests \(Washington Post\)](#)

[Female crash test dummy can reduce injurie \(Chalmers\)](#)



## Related research

Scenarios tested using female body-type crash test dummies in the driver's seat

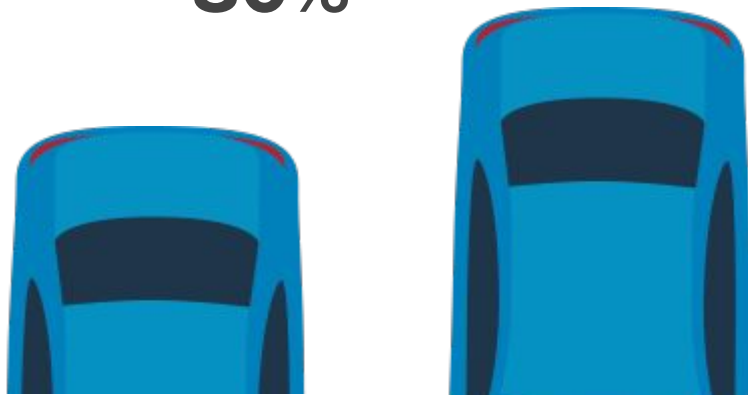
**1 of 3**

As of 2015, the number of women working in the motor vehicle manufacturing industry

**26.7%**

As of 2015, the number of car buying decisions influenced by women

**80%**



---

SOURCE

[Women in Cars - A Mega Trend for the Automotive Industry \(Frost & Sullivan\)](#)

[Vehicle safety ratings \(National Highway Traffic Safety Administration\)](#)

[Labor Force Statistics \(United States Department of Labor\)](#)

---

**We can't remove human  
perception from the loop.**

**And we can't be gripped by  
inaction either.**

---

---

The inequity demonstrated in these examples may feel overwhelming, perhaps even a little disheartening.

**But we're in the right place at the right time and in the right industry to do something about it.**

---