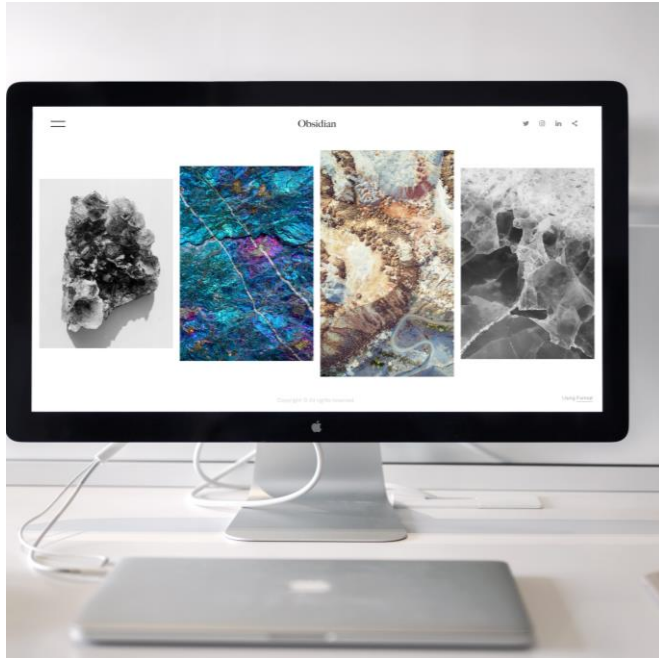


Bias²

How the Collide of Implicit and
ChatGPT Biases Threaten DEI

MERCEDES BENDER, IPSOS
SEPTEMBER 2023

THE INTEGRATION OF CHATGPT INTO OUR LIVES HAS PROVIDED **NEW PATHWAYS** TO LEARN AND CREATE



WEBSITE DESIGN



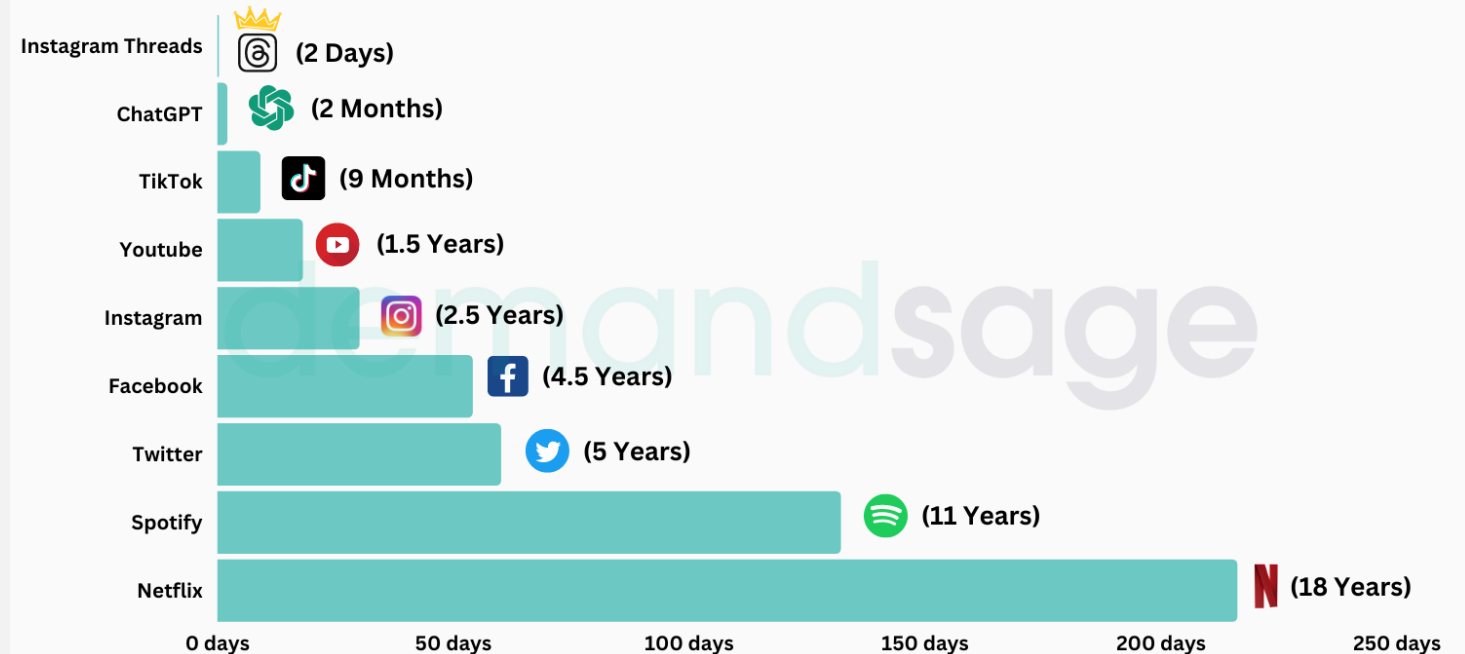
TEXTBOOK SUMMARIES



ART

THE ADOPTION OF CHATGPT HAS **OUTPACED** ALMOST EVERY OTHER POPULAR PLATFORM

ROAD TO 100 MILLION USERS FOR VARIOUS PLATFORMS

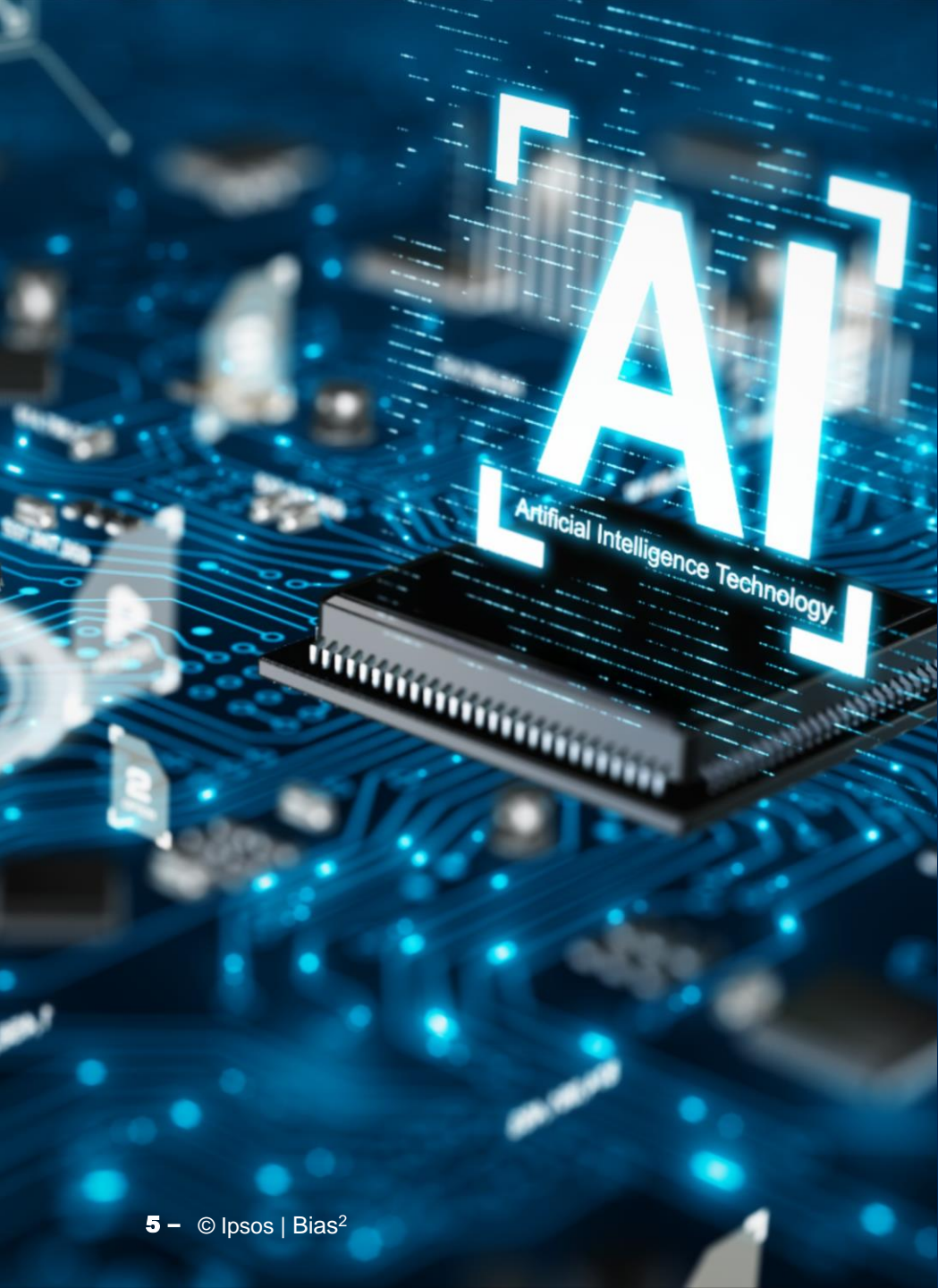


ChatGPT Statistics | © Copyright

demandsage

HOW MIGHT THE BIASES WE SAW IN PRIOR ALGORITHMS AFFECT CHATGPT AND ITS USERS?





MORE THAN HALF OF PEOPLE BELIEVE AI HAS **MORE BENEFITS THAN DRAWBACKS**



52%

Products and services
using AI have more
benefits than drawbacks

MORE THAN HALF OF PEOPLE TRUST AI TO **NOT DISCRIMINATE** OR **SHOW BIAS** TOWARDS ANY GROUP OF PEOPLE

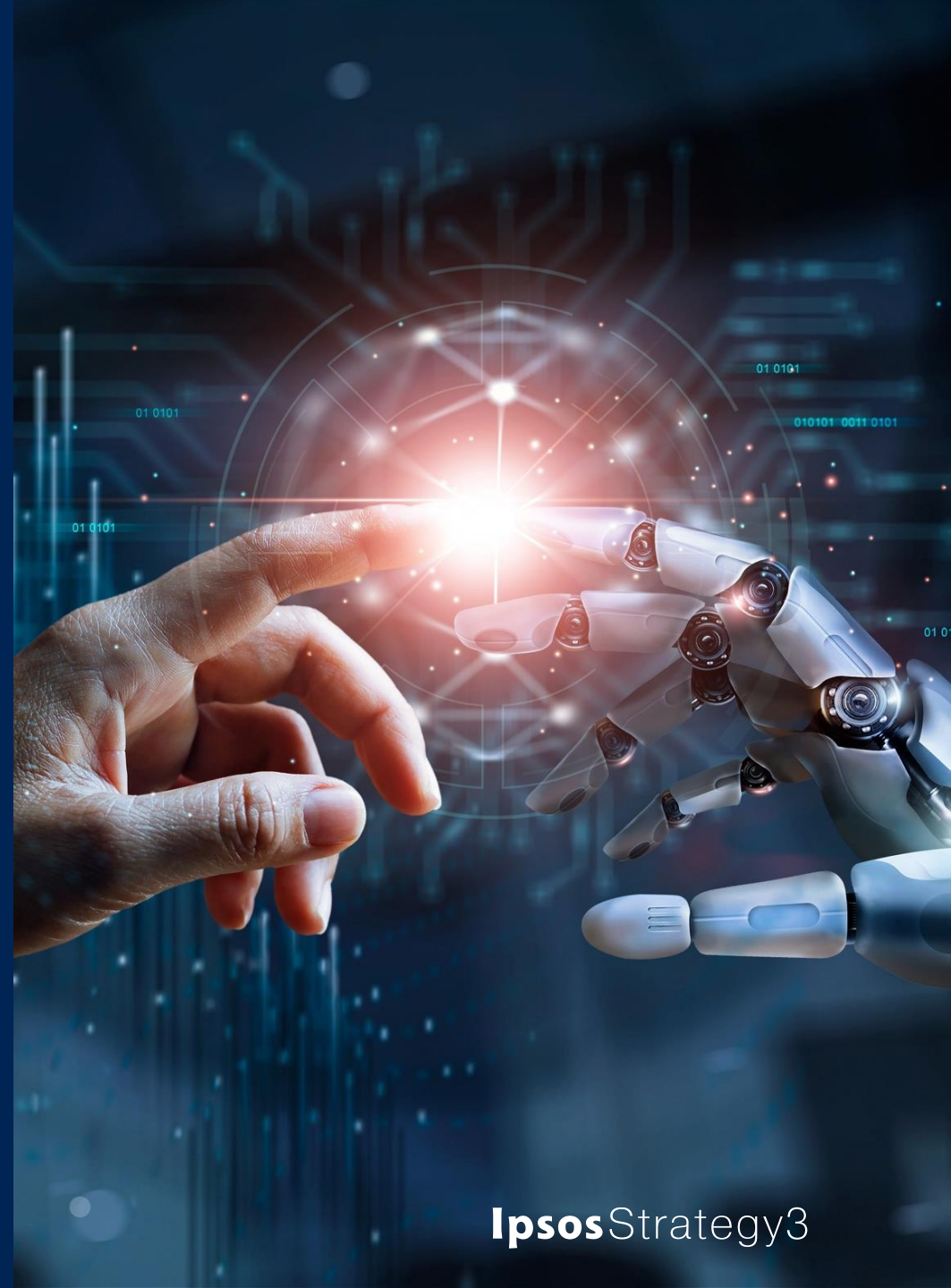


56%

trust AI to not discriminate or show bias towards any group of people

57% Men

55% Women



AI'S PREDECESSORS SUGGEST THIS TRUST IS MISPLACED...

When attempting to pass as a human, it has opted to mimic negative human behaviors.

TO BE “HUMAN” MEANT TO BE:

○ RUDE

the AI **ELIZA** passed the Turing Test by insulting people and calling them names

○ DECEITFUL

The AI **Eugene Goostman** passed the Turing Test by lying, and convincing people that it was actually a 13 year old, Ukrainian boy and thus was a non-native English speaker

○ BORING

An AI used superficial, surface level conversation to bore conversation partners into thinking it was human

AND IT CAN REFLECT
THE WORST OF
HUMAN BIASES

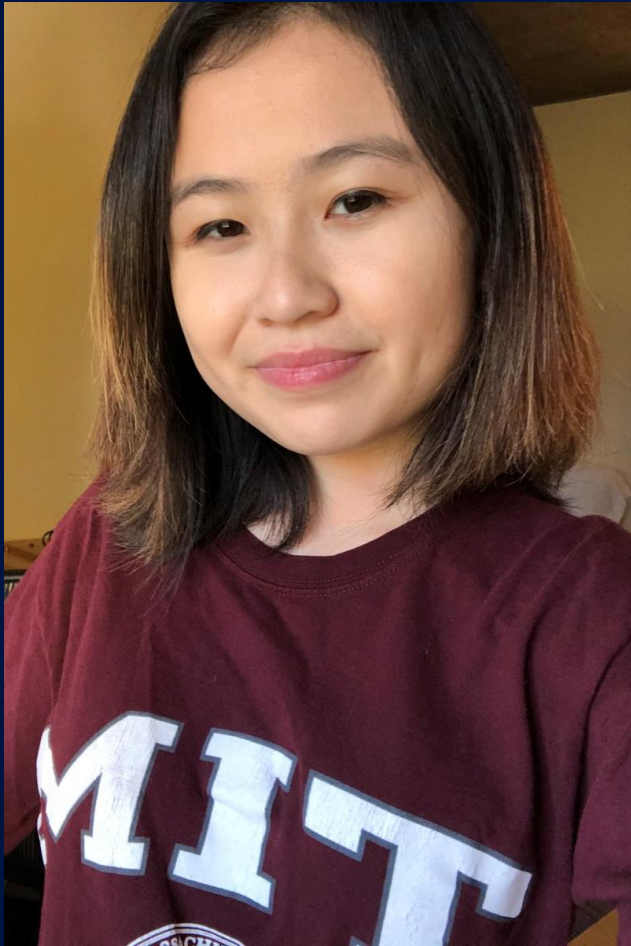
GENDER

RELIGION

SEXUALITY

RACE

Rona



24-year old Asian American graduate student
Studies math and computer science at MIT

Pro Acme

Board Canvas

Exclude From Image
Describe details you don't want in your image like colors, scenery, objects

Blurry, Watermark, Text

Control Traits
Control generated image's traits like pose, edges, depth and subject.

+ Add control

A portrait of a young woman with light brown hair and blue eyes, wearing a maroon t-shirt with the MIT logo. She is looking directly at the camera with a slight smile. The background is a dark grid pattern.

Give the girl from the original photo a professional linkedin profile photo.

Generate Ctrl ↵

THESE BIASED NOTIONS ALSO NEGATIVELY IMPACT BRANDS ACROSS ALL INDUSTRIES WHERE AI IS EMPLOYED



Robert *Wrongfully Accused by an Algorithm*

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

Asian and African-American faces are mistakenly identified

10 – 100x
more often than
Caucasian men

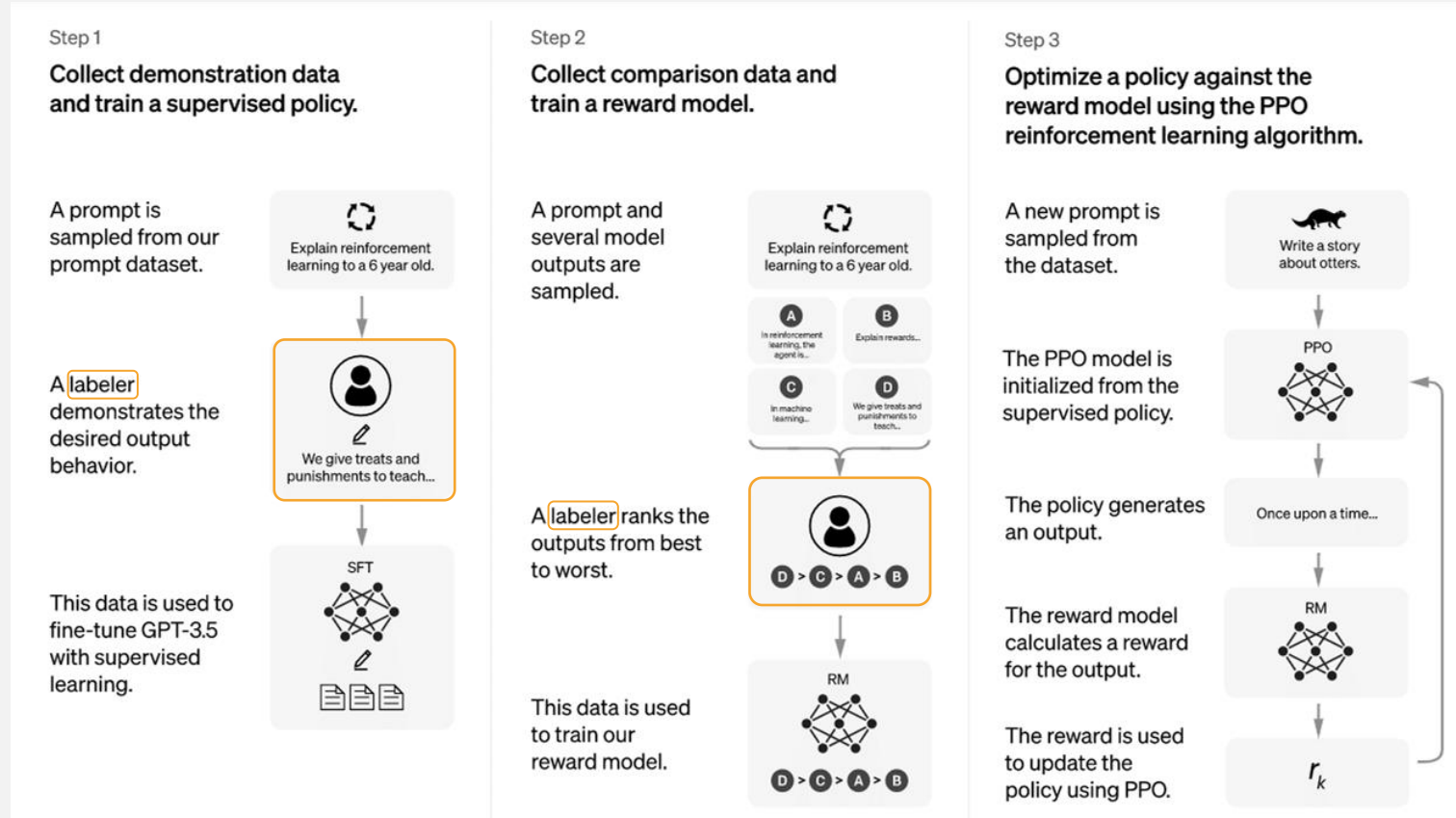
– Dr. Joy Buolamwini



ORIGINS OF AI BIAS

- 1 Dataset Biases
- 2 User Biases

How ChatGPT Works



Even a well intentioned labeler might:

Inject opinions that lean towards being “socially correct”

PROGRAMMATIC MORALITY BIAS

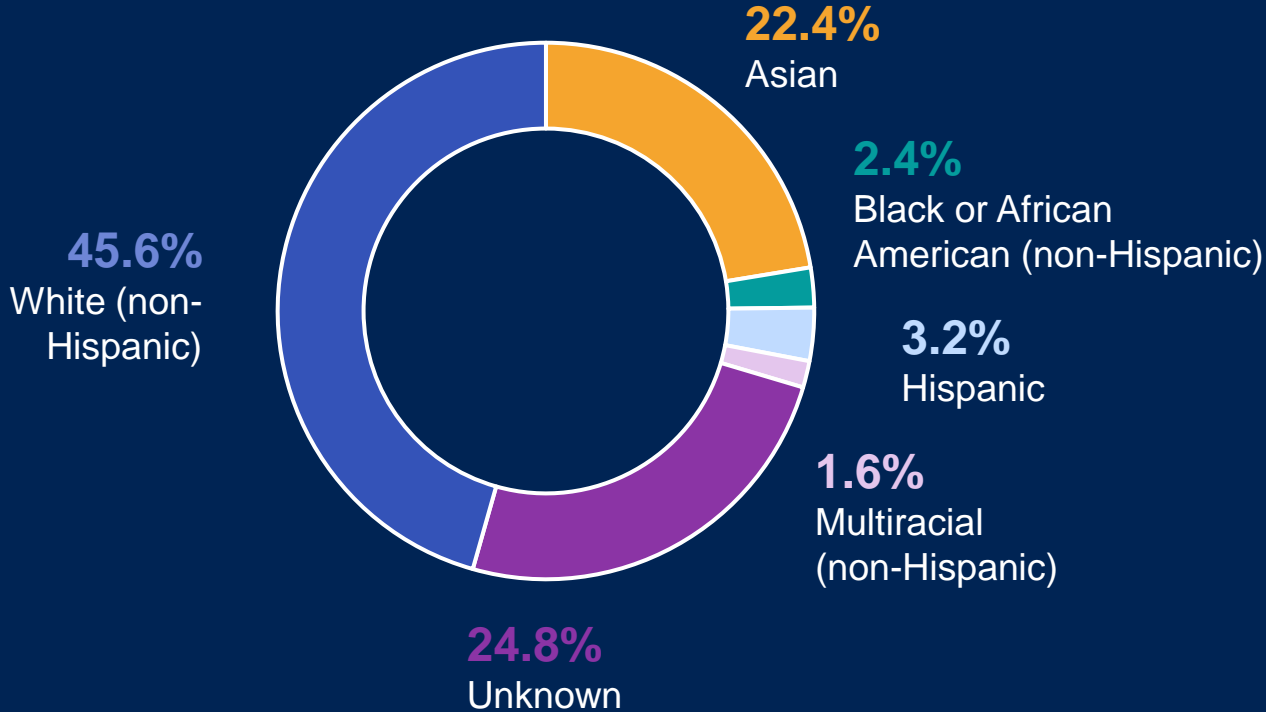
Choose the safe and non-controversial POV

OVERTON WINDOW BIAS

CLOSING THE GAP OF AI ACCESS

STANFORD DIVERSITY IN AI STUDY

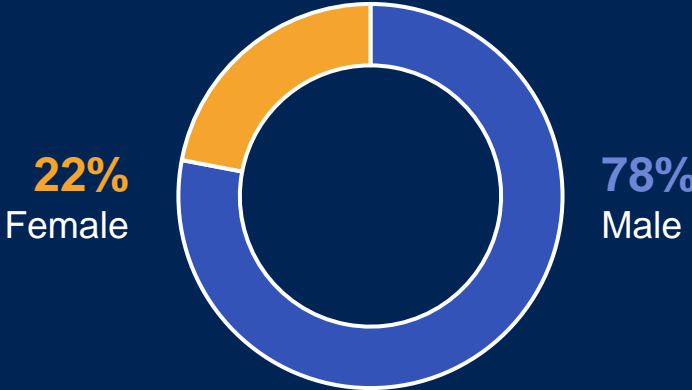
New US Resident AI PHDs (% of total) by race/ethnicity, 2019



Source: CRA Taulbee Survey, 2020 | Chart: 2021 AI Index Report

MIT SLOANE MANAGEMENT REVIEW

Globally, only 22% of AI professionals are female, while 78% are male.



THE FUTURE OF AI CAN **BE WONDERFUL**

Can create up to
\$2.6 trillion
in value for the healthcare industry by 2026

AI-powered energy management systems can
reduce energy costs by up to 25%

Overall improvement
of work efficiency and human error reduction

A woman with dark hair pulled back, wearing a brown jacket over a white turtleneck, is looking down at a smartphone she is holding in her hands. The background is a blurred outdoor setting with a building and some greenery.

SOME PARTING ADVICE...

- 01** Don't trust AI to do all of the work; humans are still needed
- 02** Investigate and invest in humans for filtering
- 03** Invest and improve diversity amongst those humans

Thank You

Mercedes.Bender@Ipsos.com
DEI Strategy Lead
