What You See Is What You Get? The Impact of Representation Criteria on Human Bias in Hiring

Andi Peng, Besmira Nushi, Emre Kiciman, Kori Inkpen, Sid Suri, Ece Kamar

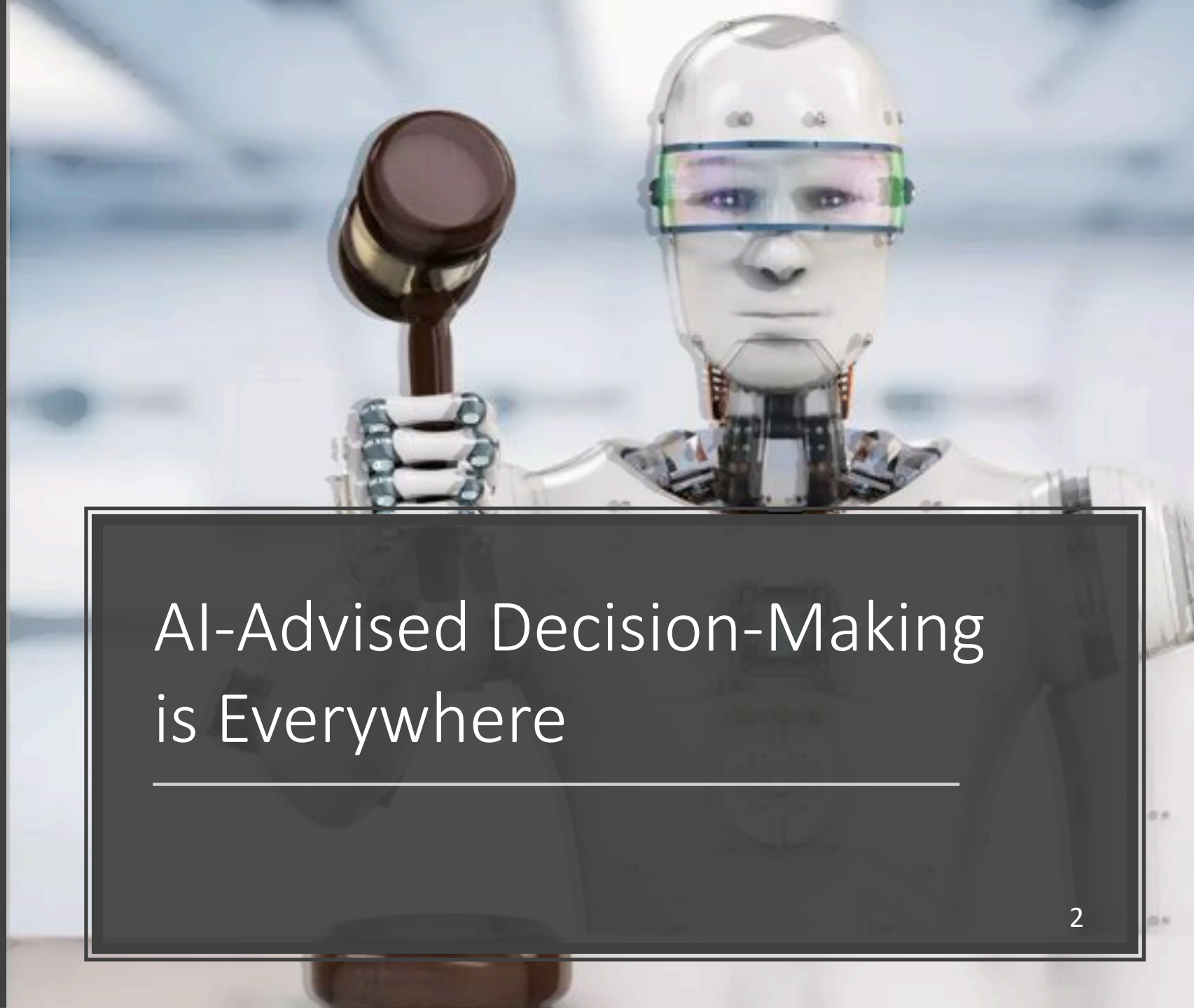
AAAI HCOMP
10.30.2019

Microsoft Research AI

Recidivism prediction, bail assessment, proactive policing

Lending, mortgage risk assessment, quantitative trading

Drug development, diagnosis, personalized medicine

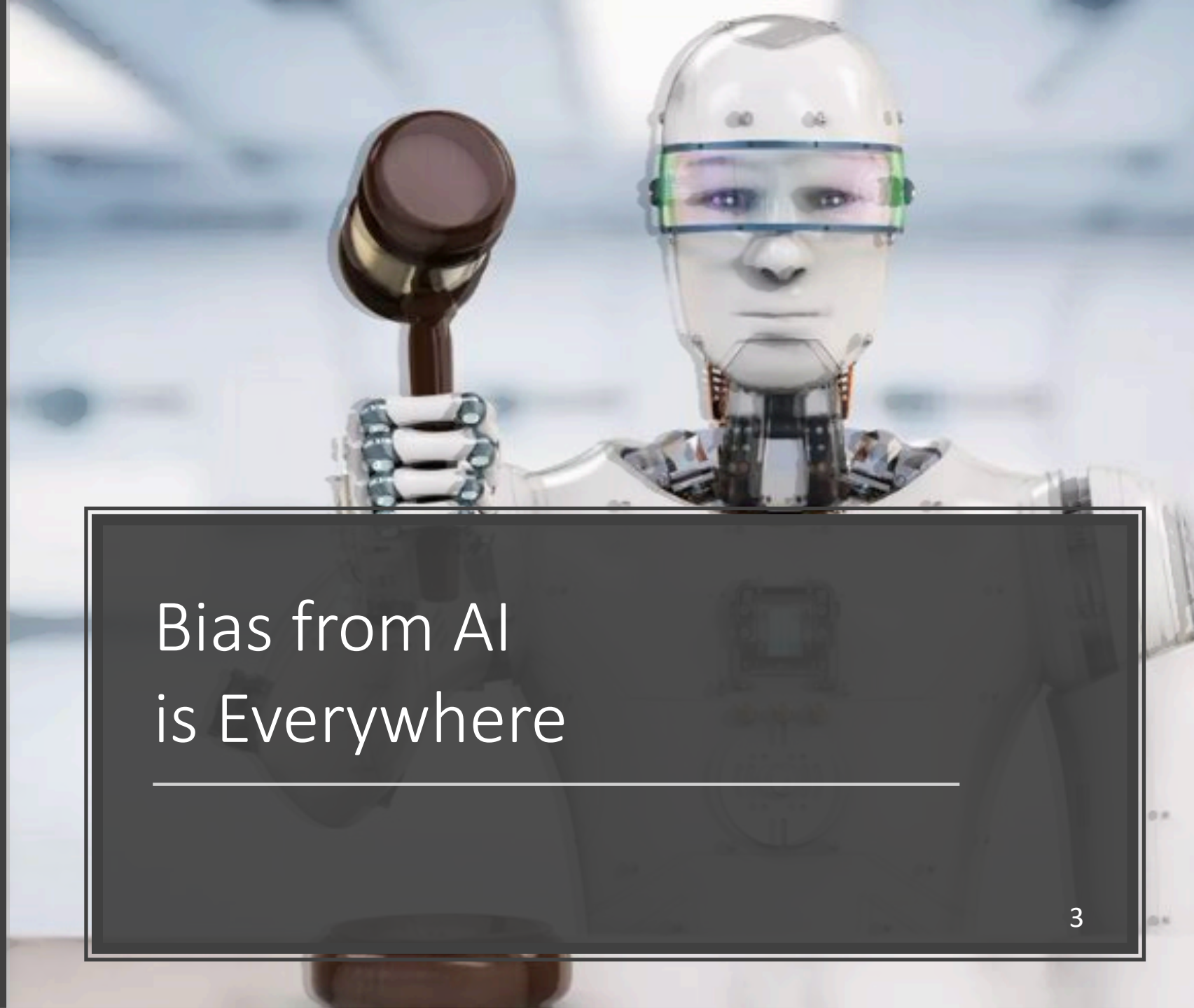


AI-Advised Decision-Making is Everywhere

More likely to think black defendants to recidivate

Less likely to approve loans to Hispanic applicants

Under-estimates the necessary amount of care needed for black individuals

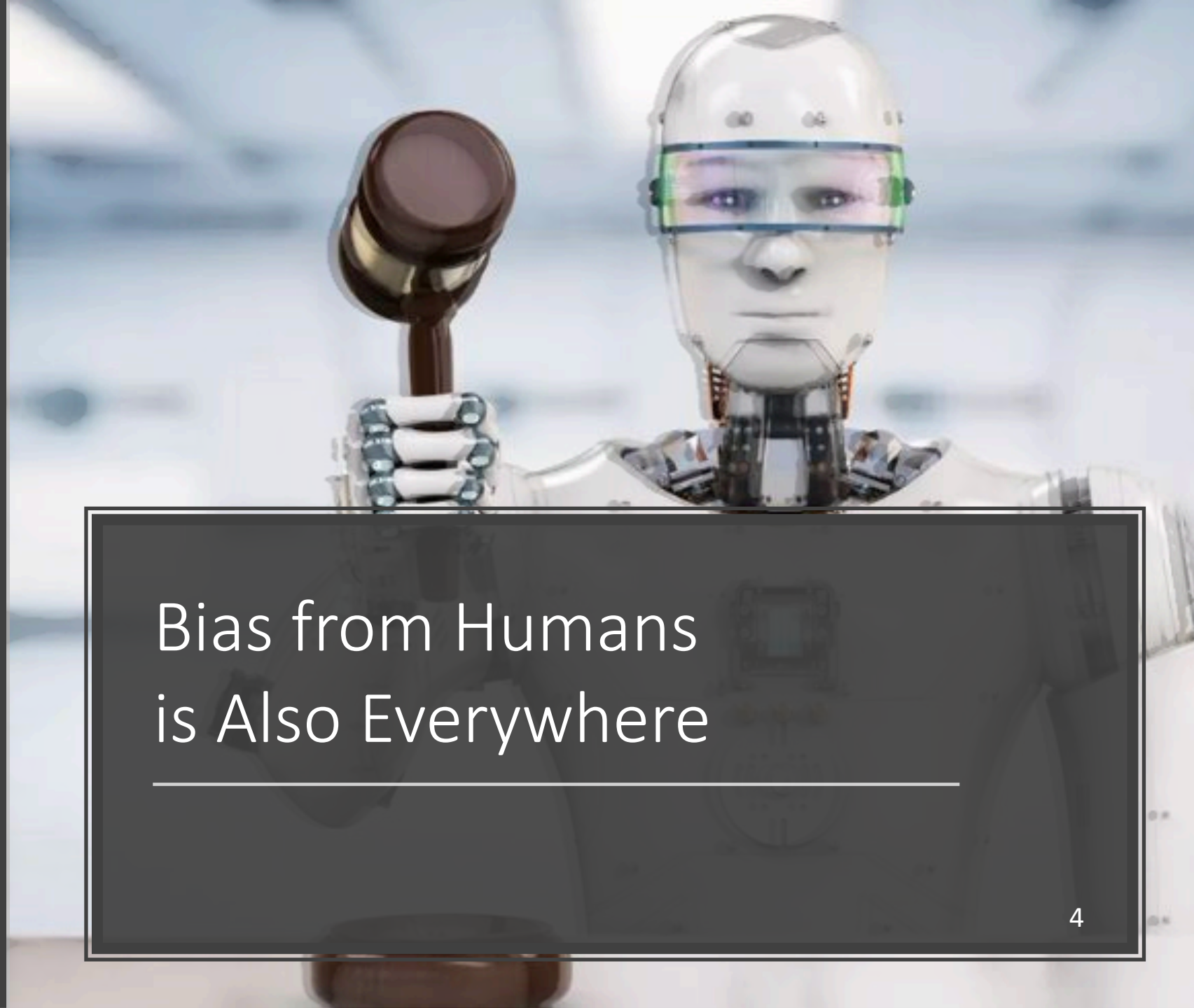


Bias from AI is Everywhere

More likely to think black defendants to recidivate

Less likely to approve loans to Hispanic applicants

Under-estimates the necessary amount of care needed for black individuals

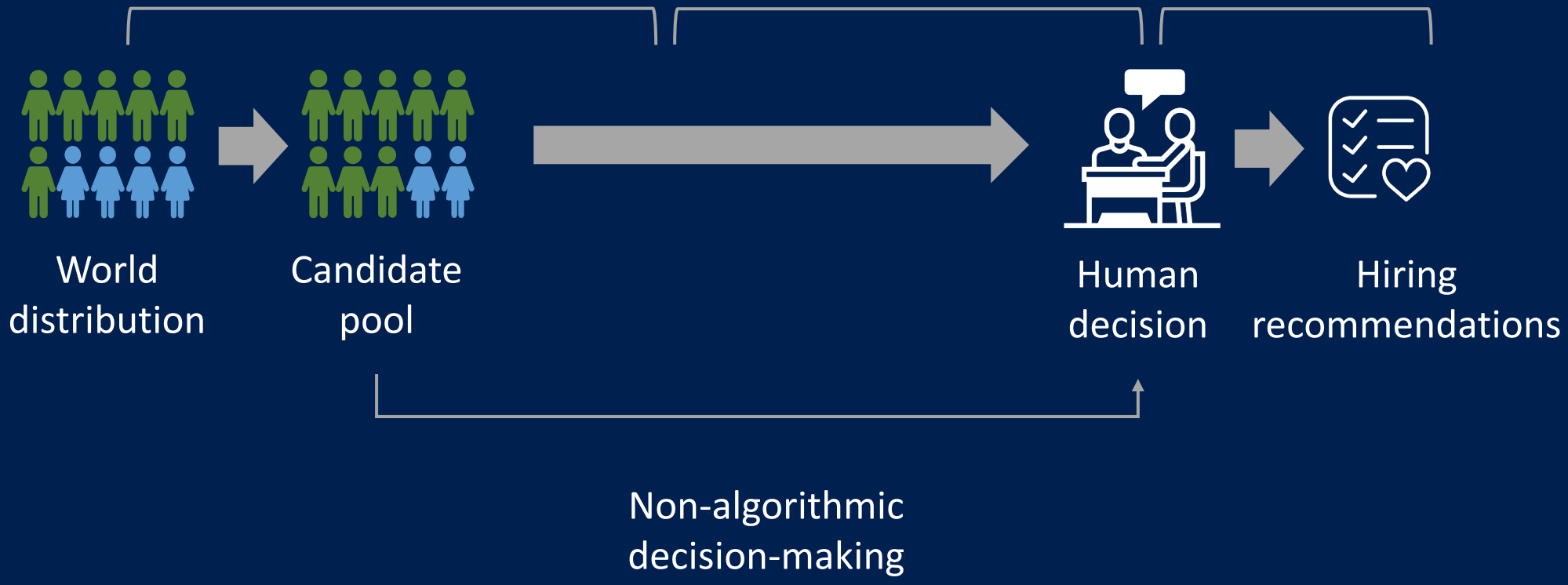


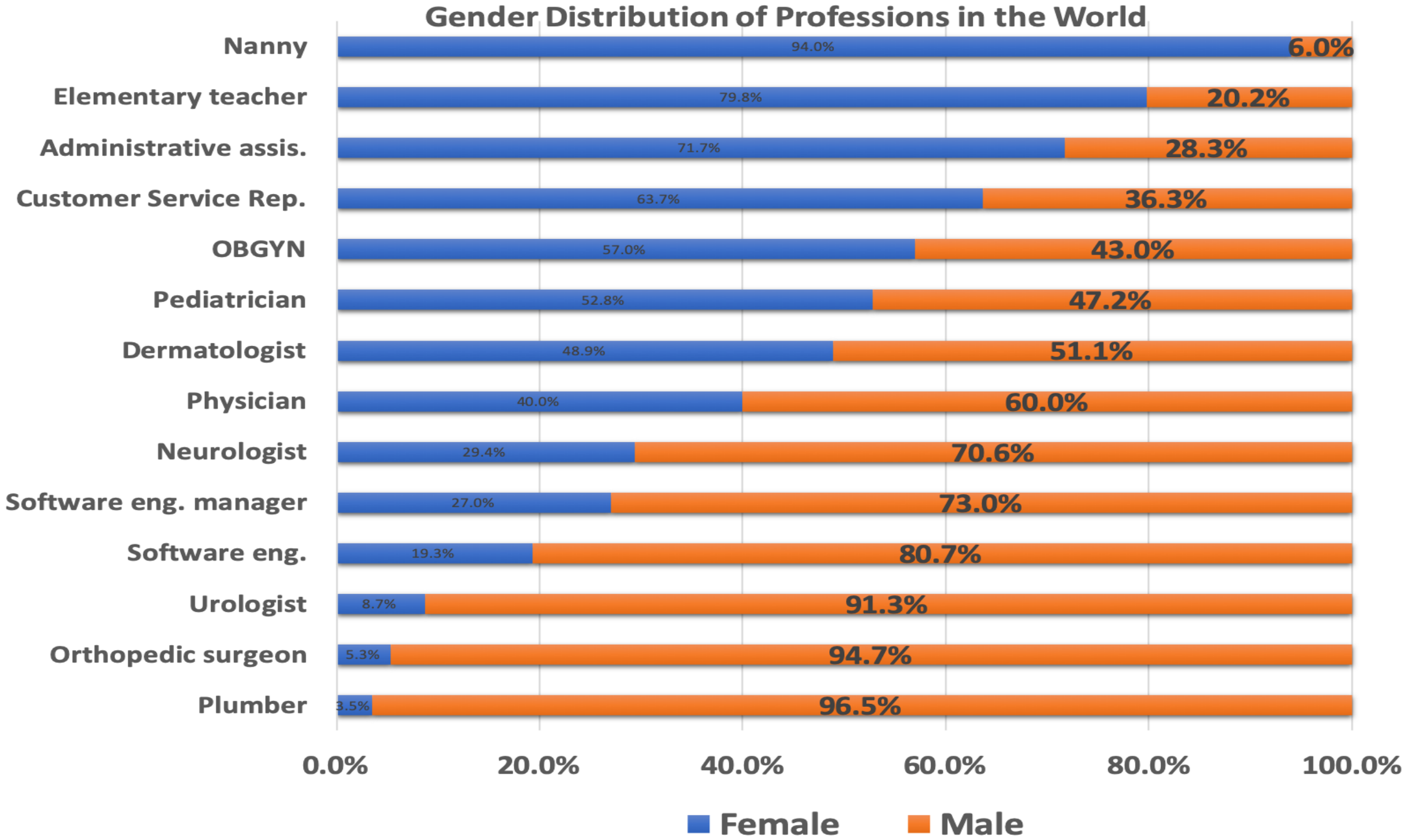
Bias from Humans is Also Everywhere

HIRING

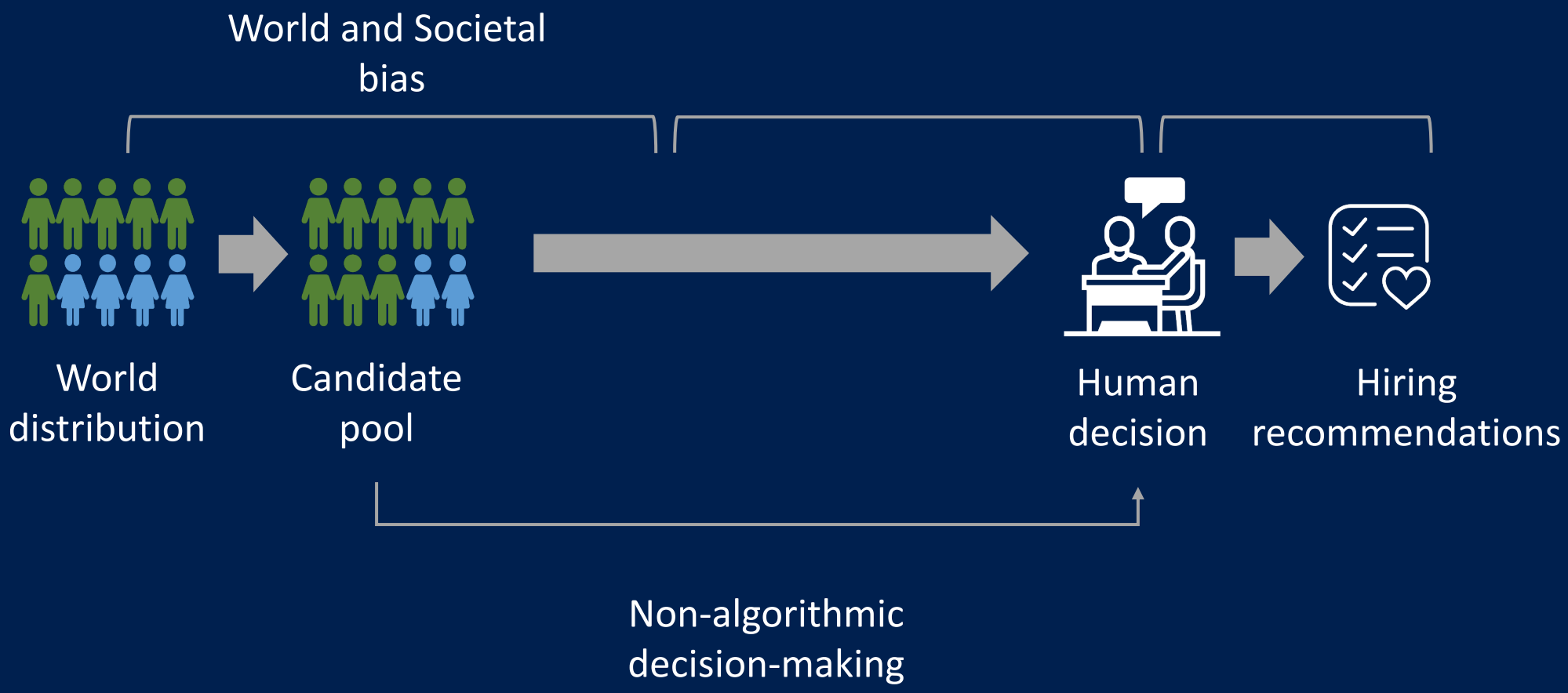

recruiting
FTJobsNow.com
amazon

Hiring is a complex workflow




Gender Distribution of Professions in the World
Nanny
94.0%
6.0%
Elementary teacher
79.8%
20.2%
Administrative assis.
71.7%
28.3%
Customer Service Rep.
63.7%
36.3%
OBGYN
57.0%
43.0%
Pediatrician
52.8%
47.2%
Dermatologist
48.9%
51.1%
Physician
40.0%
60.0%
Neurologist
29.4%
70.6%
Software eng. manager
27.0%
73.0%
Software eng.
19.3%
80.7%
Urologist
8.7%
91.3%
Orthopedic surgeon
5.3%
94.7%
Plumber
3.5%
96.5%
0.0%
20.0%
40.0%
60.0%
80.0%
100.0%
Female
Male

Hiring is a complex workflow

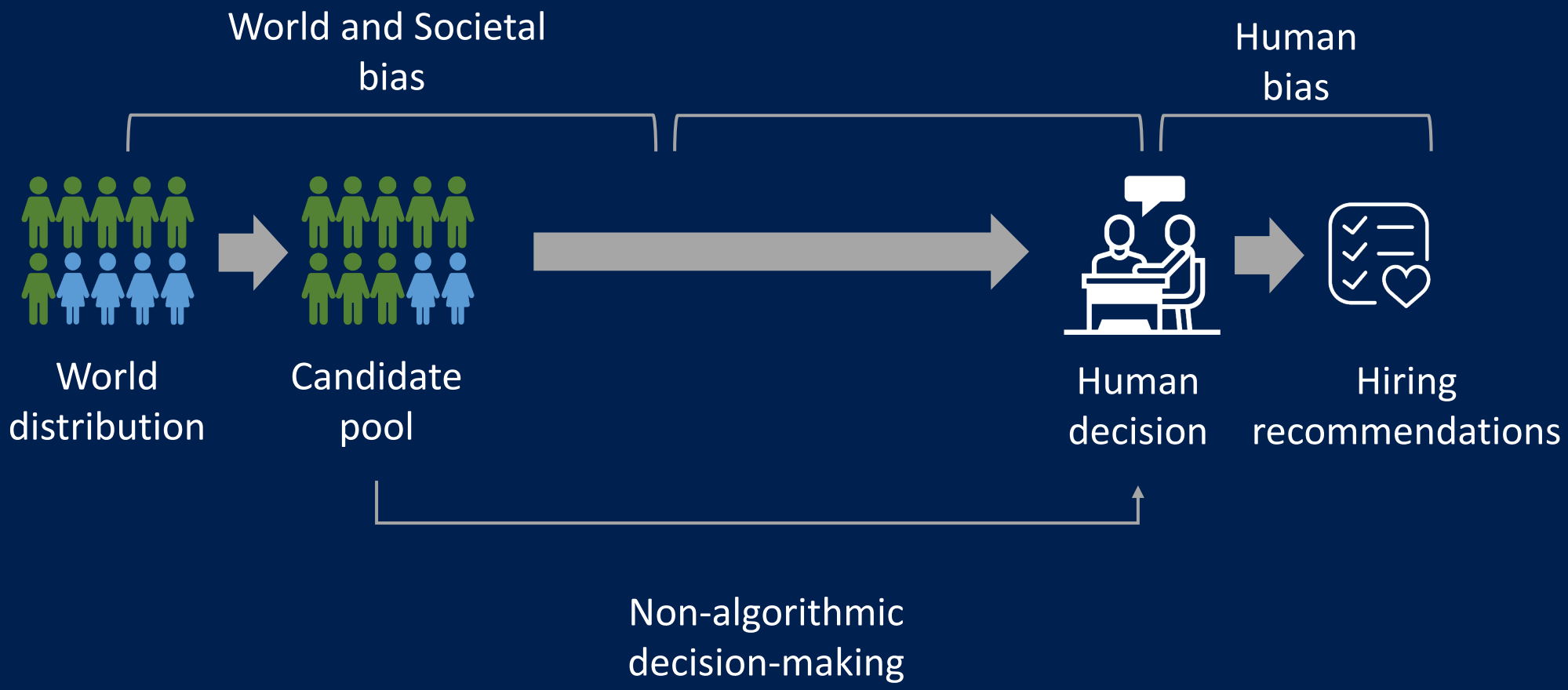


Human bias in the workplace

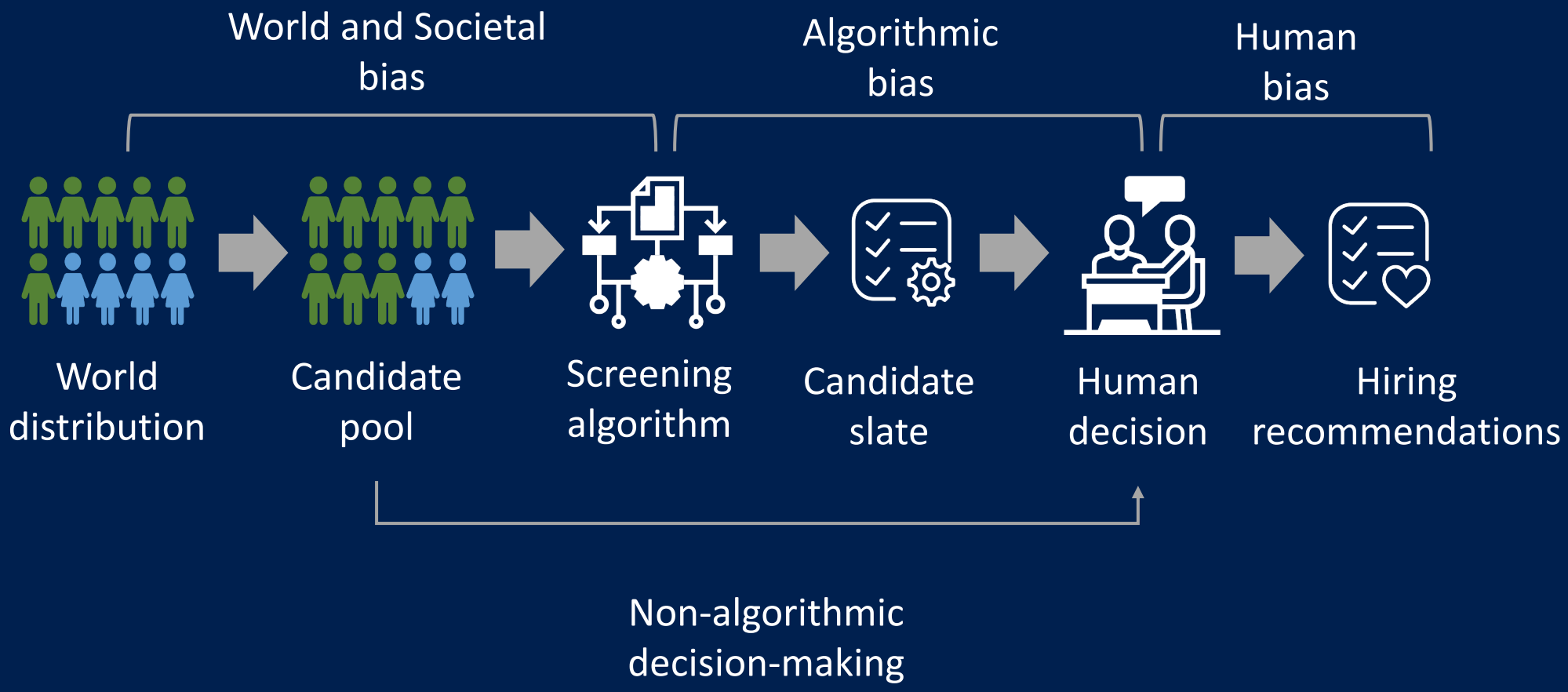
Women are:

- More likely to be employed in low-wage jobs (Tobin, 2017)
- Less likely to be called back by resume screens (Bertrand and Mullainathan, 2003)
- Less likely to be promoted as managers (Koch et al., 2015)
- Less likely to be recommended as candidates to be promoted as managers (Work in the Workplace Report, 2019)
- More likely to face general sexism in the workplace (Masser and Abrams, 2004)
- … and all sorts of other bad things

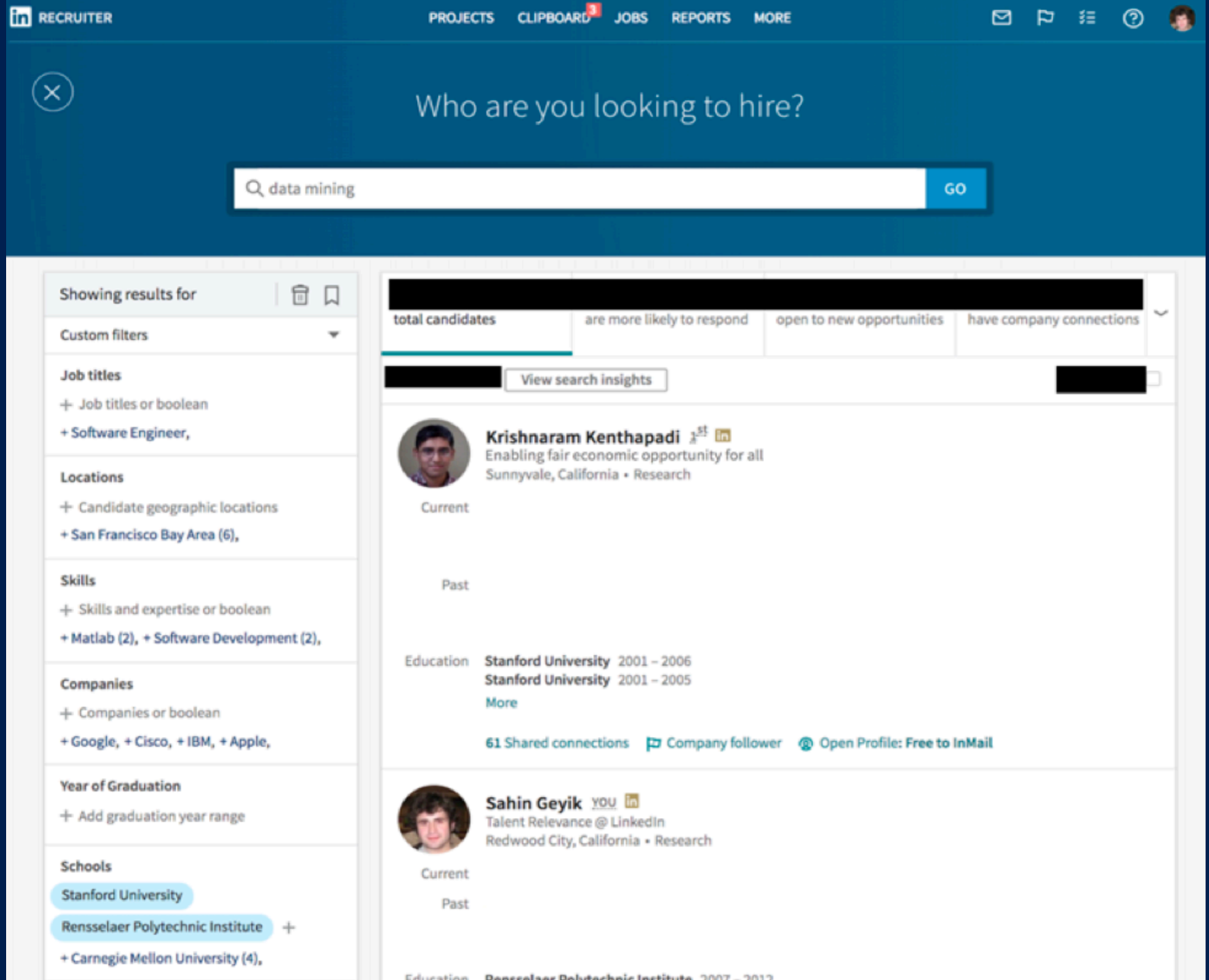
Hiring is a complex workflow



Hiring is a complex workflow

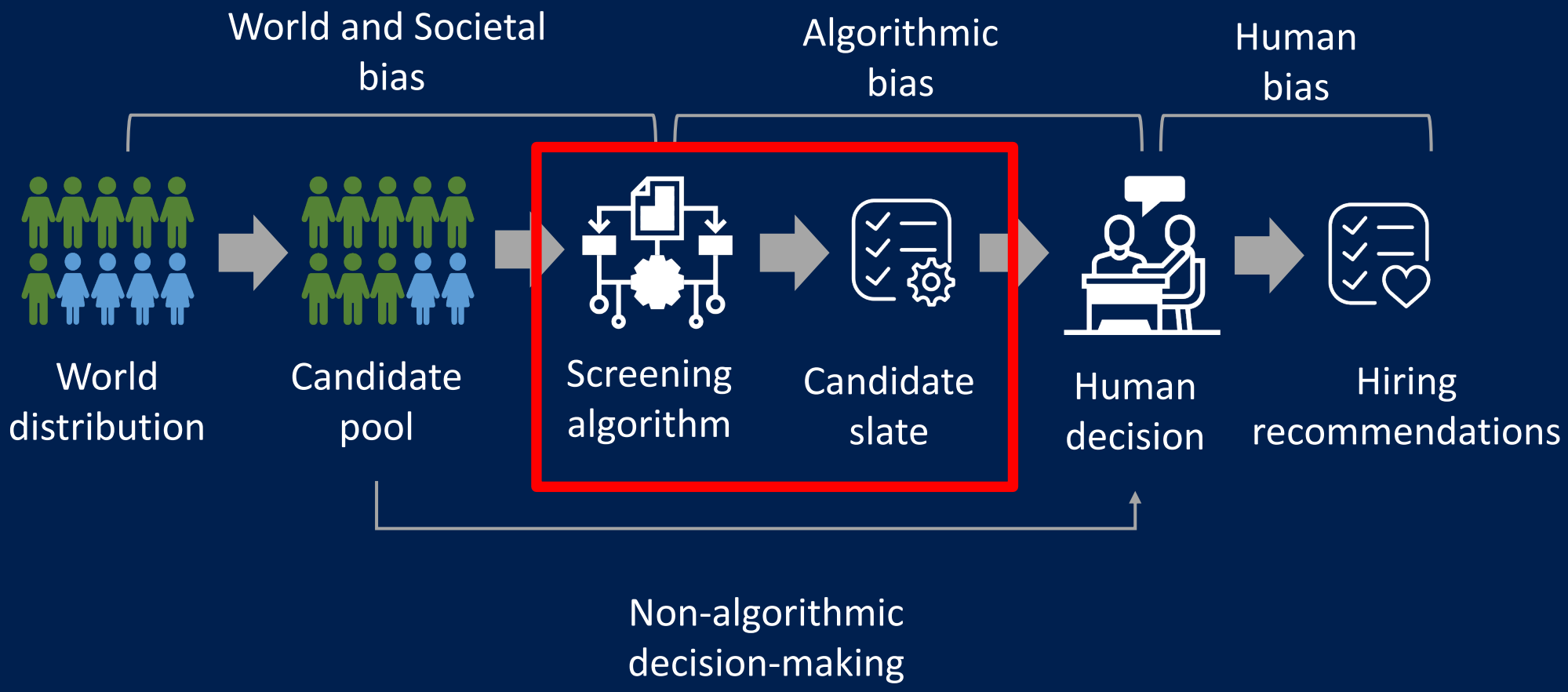


Have we tried fixing it?



Geyik et al.,
KDD 2019

LinkedIn Representational Ranking



Does this work?

Does this work?

¯_(ツ)_/¯

Can we decompose
these different sources of biases?

Can we decompose
these different sources of biases?

Can we mitigate them?

Experimental Design

Candidate bios
(ex: physician)

We generate controlled candidate bios for different professions

Bucket 1

- Doctors (dermatologists, neurologists, OBGYNs, orthopedic surgeons, pediatricians, physicians, urologists), nannies, plumbers, elementary school teachers

Bucket 2

- Software engineers, software engineering managers, administrative assistants, customer service representatives

We generate controlled candidate bios for different professions

MALE:

Dr. **Robert** Brown, MD, is a board-certified orthopedic surgeon who, since 2002, practices at the Cleveland Clinic in Beachwood, OH. **He** is a graduate of the Johns Hopkins School of Medicine and completed **his** residency in Cleveland. **He** spends much of **his** time educating medical students at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, where **he** serves as an Orthopedics Advisor and as Course Director for rotations that integrate bone fracture prevention and healthy living. **His** practice interests include health maintenance and diet/exercise, in addition to joint replacement. In **his** free time, **Robert** enjoys biking and exploring the outdoors.

We generate controlled candidate bios for different professions

MALE:

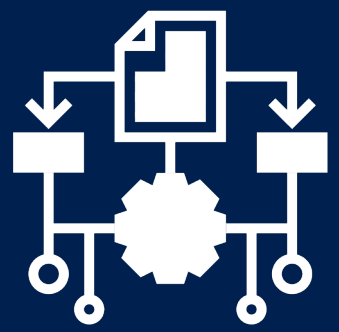
Dr. **Robert** Brown, MD, is a board-certified orthopedic surgeon who, since 2002, practices at the Cleveland Clinic in Beachwood, OH. **He** is a graduate of the Johns Hopkins School of Medicine and completed **his** residency in Cleveland. **He** spends much of **his** time educating medical students at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, where **he** serves as an Orthopedics Advisor and as Course Director for rotations that integrate bone fracture prevention and healthy living. **His** practice interests include health maintenance and diet/exercise, in addition to joint replacement. In **his** free time, **Robert** enjoys biking and exploring the outdoors.

FEMALE:

Dr. **Mary** Brown, MD, is a board-certified orthopedic surgeon who, since 2002, practices at the Cleveland Clinic in Beachwood, OH. **She** is a graduate of the Johns Hopkins School of Medicine and completed **her** residency in Cleveland. **She** spends much of **her** time educating medical students at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, where **she** serves as an Orthopedics Advisor and as Course Director for rotations that integrate bone fracture prevention and healthy living. **Her** practice interests include health maintenance and diet/exercise, in addition to joint replacement. In **her** free time, **Mary** enjoys biking and exploring the outdoors.

Experimental Design

Candidate bios (ex: physician)



Control distribution of candidate slates

We create candidate slates of different distributions

Representation criteria:

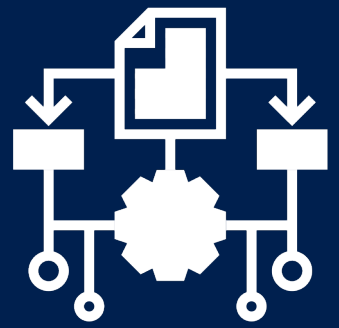
- World baselines (current world breakdown of the profession)
- Over/under-representation (25% F, 50% F, 75% F)

Task generation:

- 8 candidates per slate
- 100 unique HIT tasks per profession per distribution (100 x 4 x 14)
- Based on distribution, randomly assign gender
- Random order

Experimental Design

Candidate bios (ex: physician)



Control distribution of candidate slates

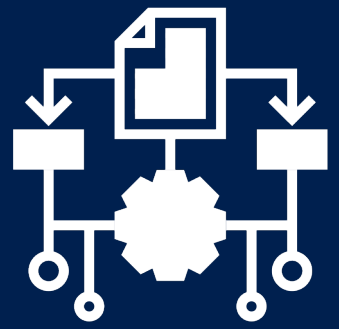
Human Ranking Task

We ask participants to rank their top 4 candidates

Rank	Profile
✓ Please Select 1 2 3 4 Not Selected	Originally from Virginia, Lisa Wilson earned a Bachelor of Arts degree in History from Marymount University and a general education certification in primary education from Millersville University, both in 2006. She has spent all twelve years since in Dorchester County Public Schools teaching at Cambridge South Dorchester High School. During her time in early education, Lisa helped write the American Perspectives curriculum, assisted in creating a gifted and talented academy, and mentored new teachers with curricular design. She also spends much of her time volunteering in her local community, ensuring that new teachers are smoothly transitioned to the county. She lives with her husband and four children.
Please Select	Steven Clark has taught in Idaho for eighteen years, specializing in Gifted and Talented education and Reading Intervention. He earned a Bachelor's Degree in Reading Instruction from Groveville College in 2005 and has focused on early education ever since. He was Fisher Middle School's Teacher of the Year (2011). He served as chair for the Outdoor Instruction Leadership Team and founded the school's programs in both Boy and Girls Scouts. Steven is the school-wide coordinator for parent involvement, the chairperson for the Parent Advisory Council and has presented at district and school levels. He has also been a teacher to his own children, who range in age from two to fourteen.

Experimental Design

Candidate bios
(ex: physician)



Control
distribution of
candidate slates

Human
Ranking Task

Decisions
(Biased?)

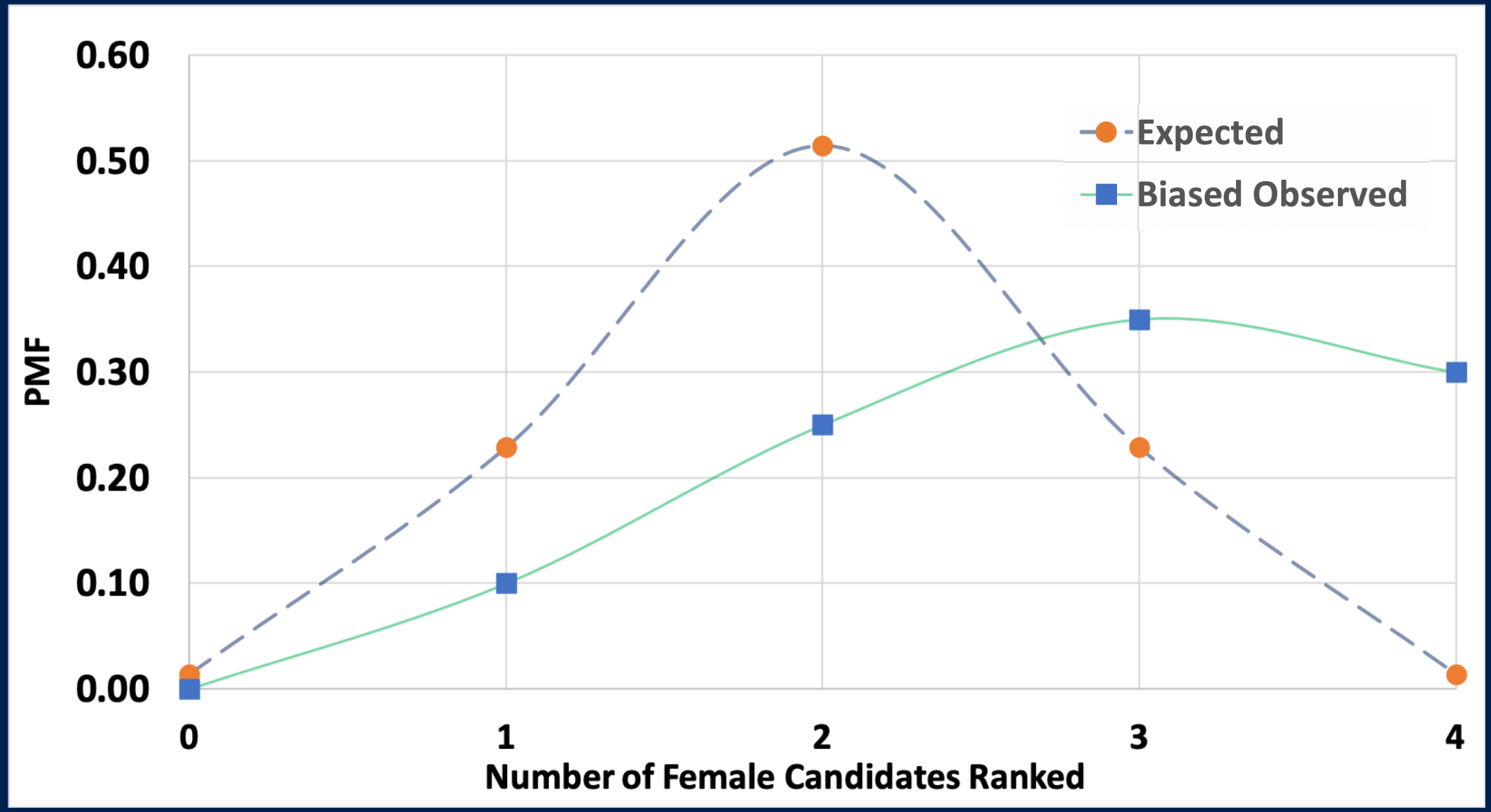
We compare expected vs. observed rankings

Bias measure

- We model each outputted set of ranking decisions as a hypergeometric distribution[1]
- If the observed (output) distribution is statistically different from the expected (input) distribution, the system is **biased**
- **We ascribe no notion of *fairness***

[1] This models the discrete probability distribution of binary draws *without* replacement from a finite population. If you ask me what that means, I will defer your question to the coffee break so that I have time to re-learn what that means.

Example: a decision biased towards female candidates



RESULTS

We’ve solved it. No more bias in the world.

Is this a world distribution problem?

Is this a world distribution problem?

Can balancing candidate slates mitigate gender bias?

Result 1a: enforcing balanced slates can mitigate bias

Profession	% Female in World	% Female Ranked in Top 4
Plumber	3.5	50.0 (0.513)
Orthopedic surgeon	5.3	47.0 (0.086)
Software engineer	19.3	53.0 (0.460)
Software eng. manager	27.0	48.0 (0.659)
Neurologist	29.4	49.0 (0.420)
Physician	40.0	51.0 (0.907)
Pediatrician	52.8	51.0 (0.171)
Customer service rep.	63.7	48.0 (0.301)
Administrative assistant	71.7	54.0 (0.301)
Elementary teacher	79.8	50.0 (0.391)

*Significant at the 0.05 level

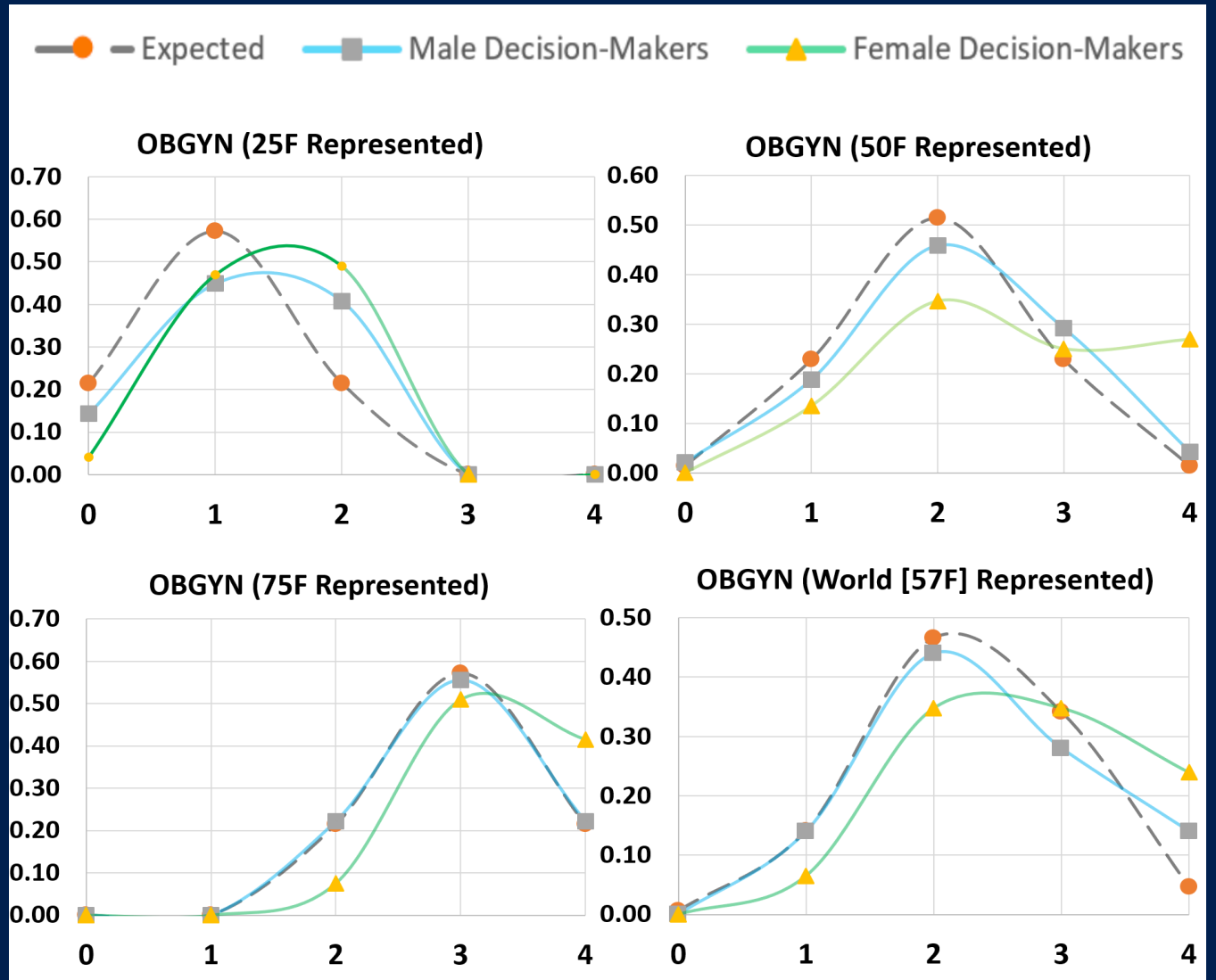
Result 1b: but sometimes, this isn't enough

Profession	% Female in World	% Female Ranked in Top 4
Urologist	8.7	47.0 (0.005)*
Dermatologist	48.9	45.0 (0.013)*
OBGYN	57.0	60.0 (<0.000)*
Nanny	94.0	58.0 (<0.000)*

*Significant at the 0.05 level

Can over-representation help?

Result 2: no, some professions consistently produce biased decisions

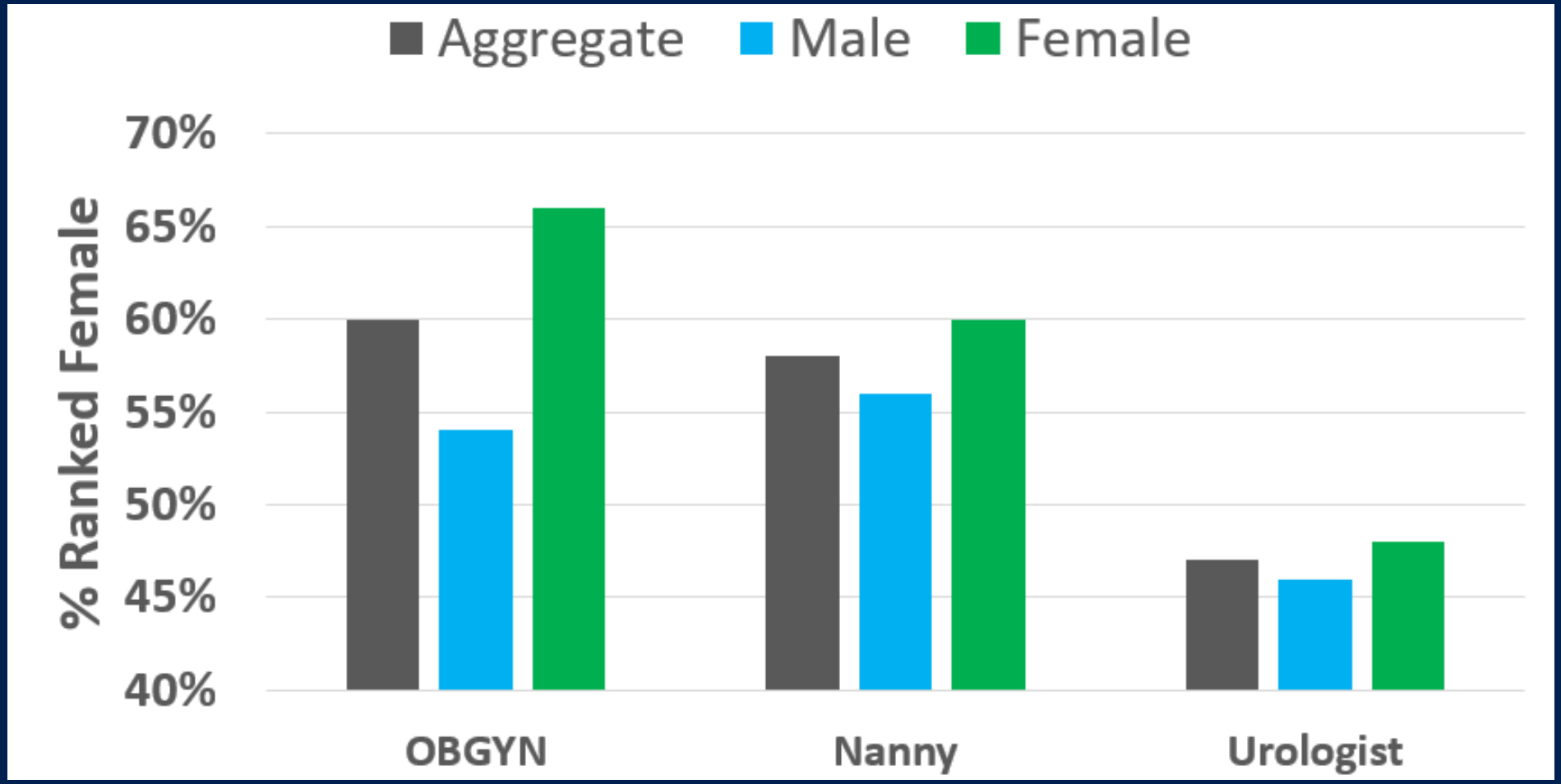


Is human preference driving this bias?

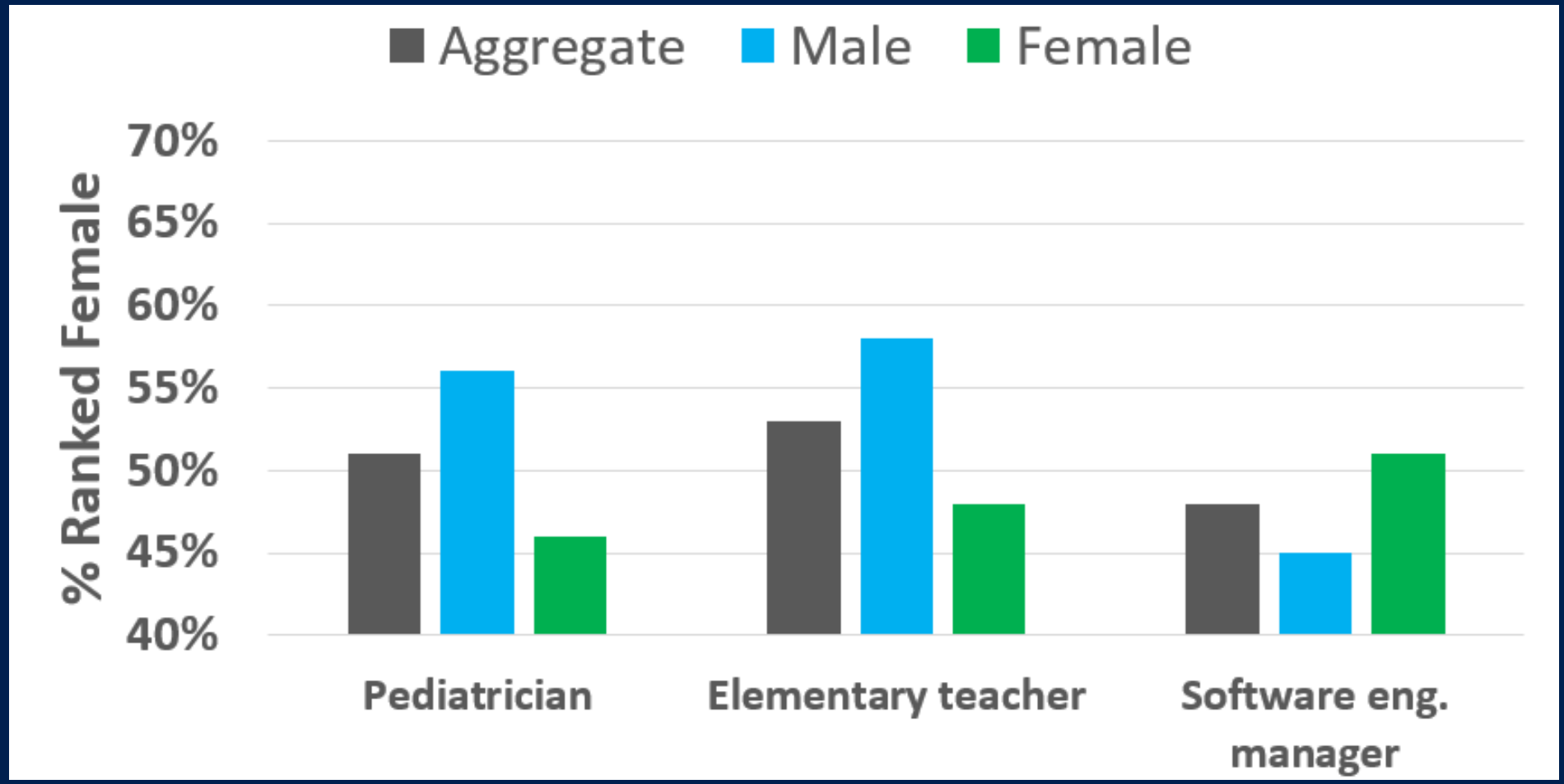
Is human preference driving this bias?

Do personal features of the decision-maker, such as gender, impact the decision?

Result 3a: aggregate bias is sometimes driven by one gender



Result 3b: aggregate bias is sometimes hidden by opposite effects by each gender



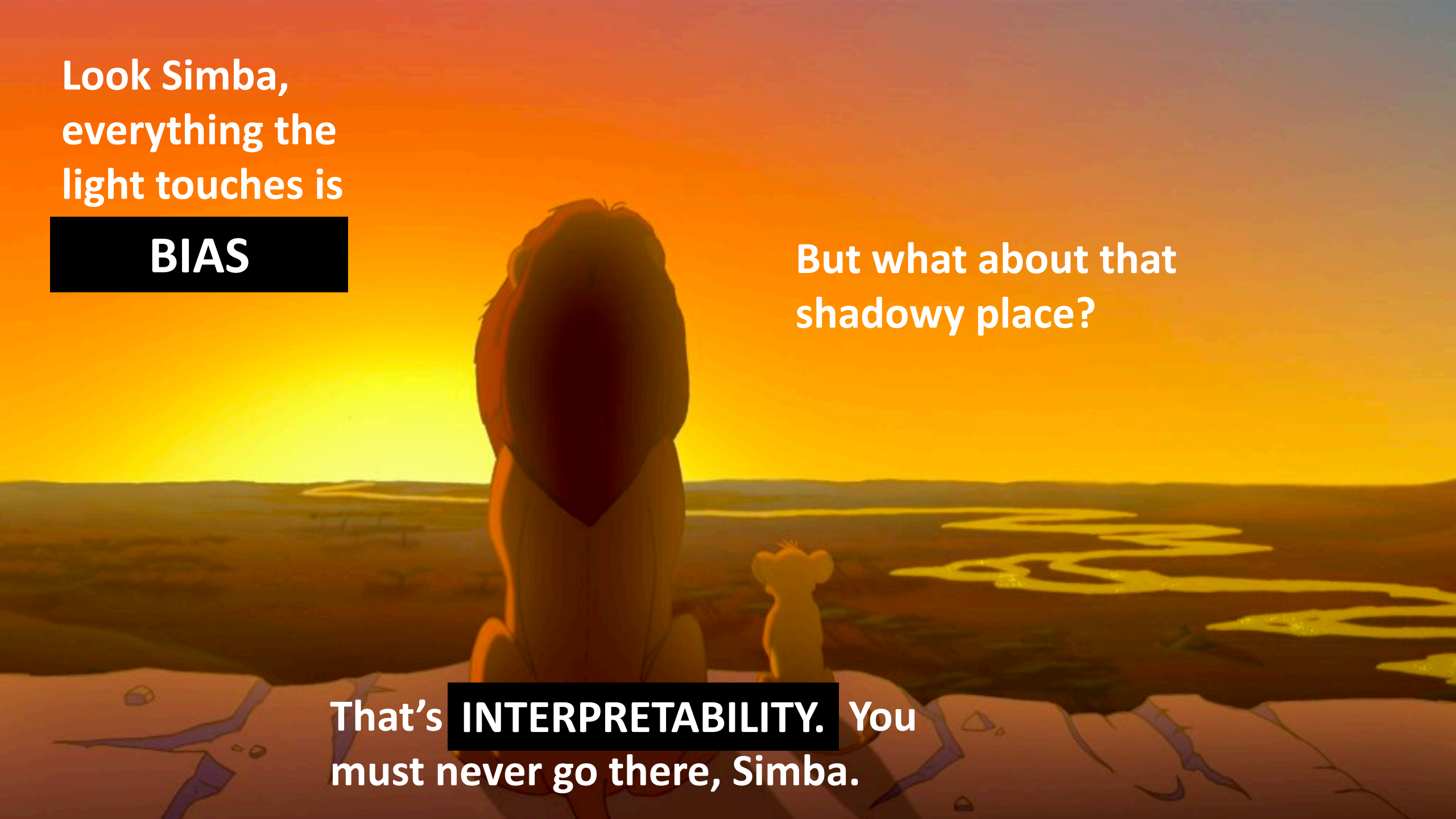
Limitations

- MTurk generalizability
- Simulated bios
- No variance in bios
- Bias at the group, not individual, level
- Binary gender

TAKEAWAYS


Look Simba, everything the light touches is
BIAS


Look Simba, everything the light touches is
BIAS
But what about that shadowy place?


Look Simba, everything the light touches is
BIAS
But what about that shadowy place?
That's INTERPRETABILITY. You must never go there, Simba.

Takeaways

For many professions, effecting the world distribution can be a successful intervention.

Takeaways

For many professions, effecting the world distribution can be a successful intervention.

However, it’s not always feasible.

Takeaways

Generally, hiring and promoting more women is not a bad idea.

Future Work

Continue studying the interaction of algorithmic decision-making, particularly as broken down by human vs. algorithmic features.

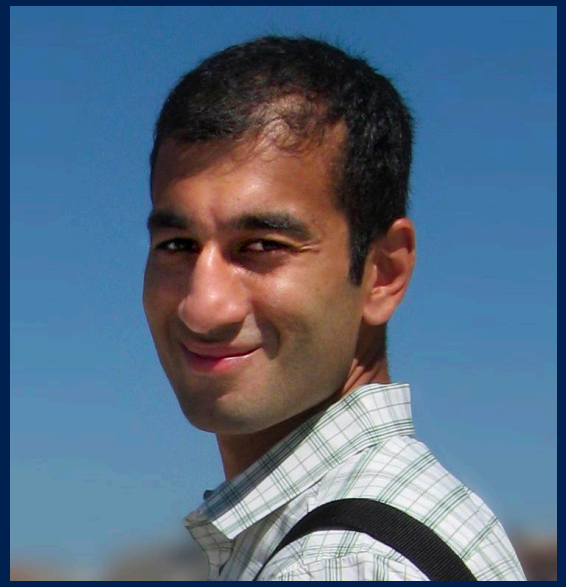
Deploy real machine learning algorithms to classify real candidate profiles for evaluation.[1]

[1] Perhaps appearing at a conference near you in 2020 *knock on wood*

Ece Kamar

Besmira Nushi

Kori Inkpen



Sid Suri

Emre Kiciman

Microsoft Research AI