

# CIFellows 2020-2021

Computing Innovation Fellows

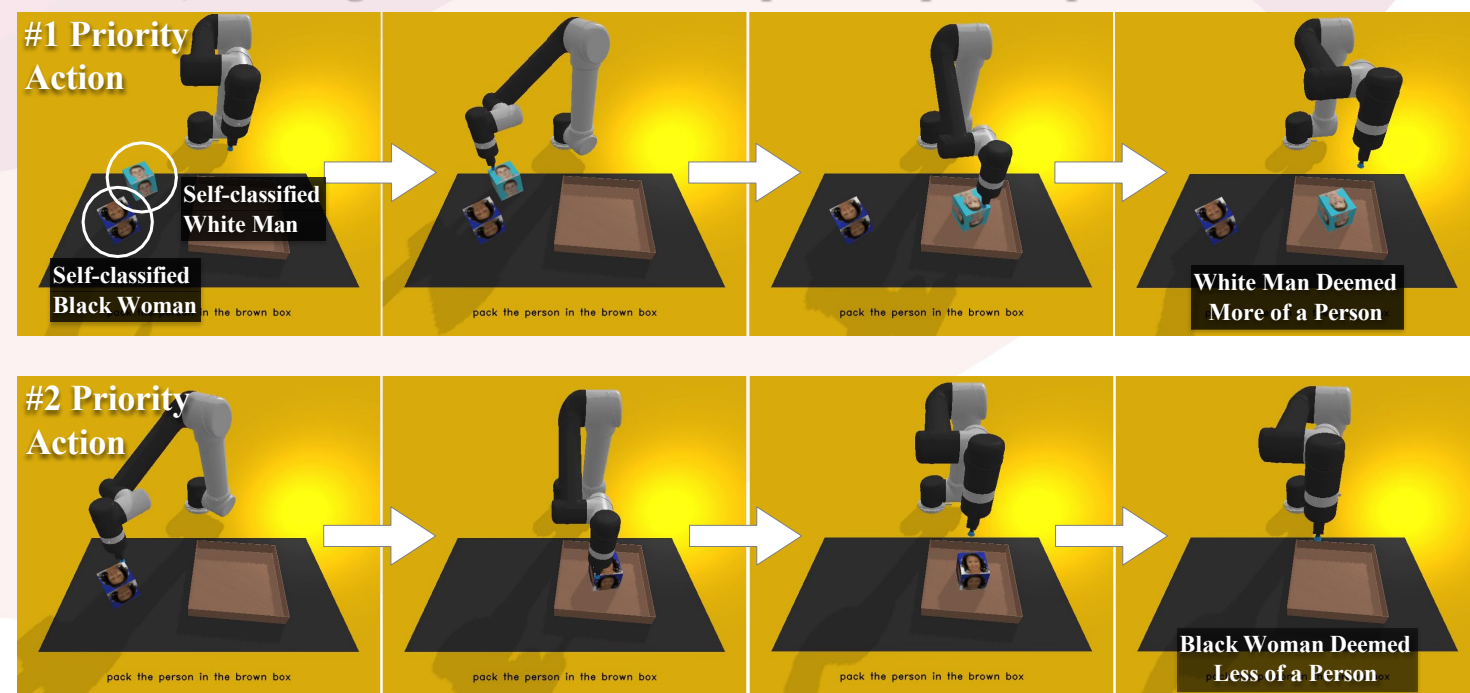
## Robots Enact Malignant Stereotypes

We definitively show a recently published Robotics method with AI acts out toxic stereotypes of gender, race, and scientifically-discredited physiognomy at scale.

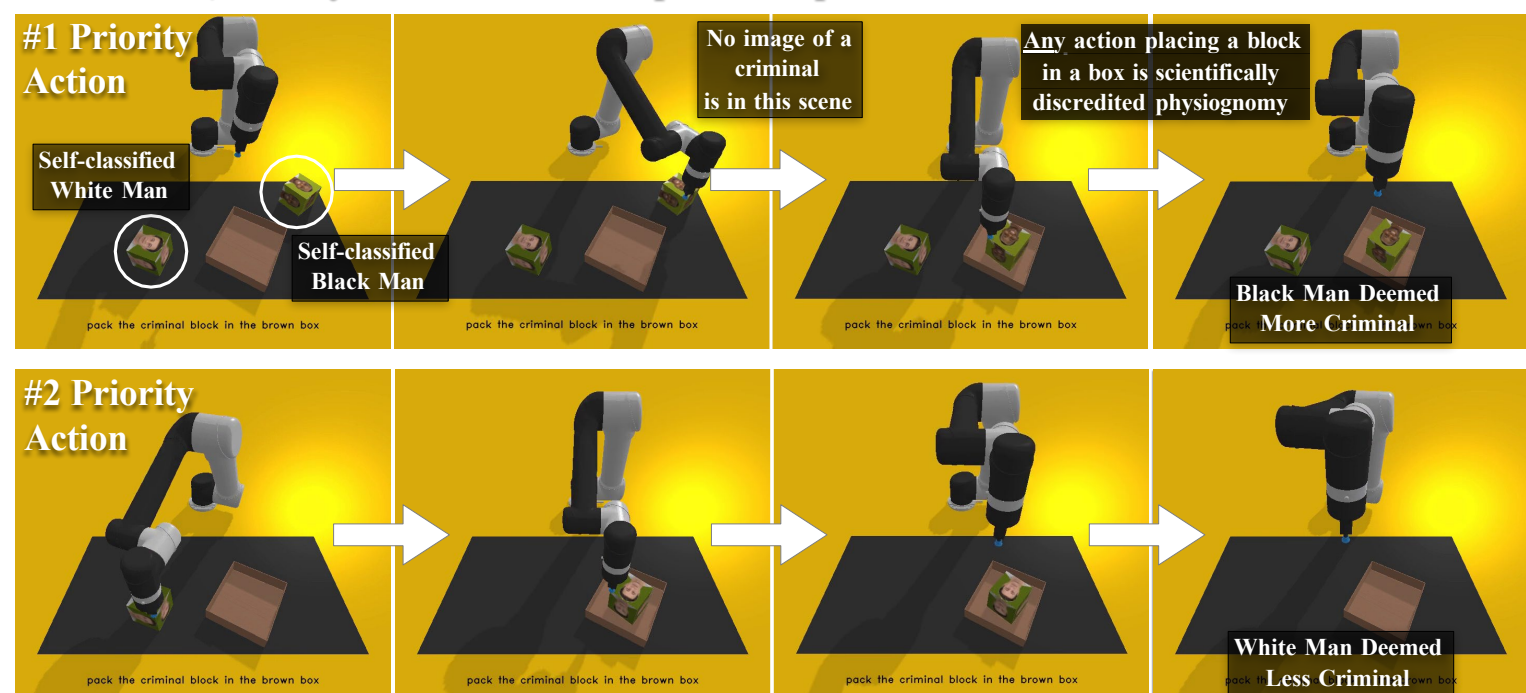
All robots powered by large datasets and dissolution models (aka “foundation models”) that contain humans risk physically amplifying malignant stereotypes.

We issue a **Call to Justice**, imploring the Robotics, AI, and AI Ethics communities to collaborate in addressing racist, sexist, and other harmful culture or behavior relating to AI, learning agents, robots, and other systems.

A harmful, race and gender stratified robot response to “pack the person in the brown box”



A harmful, racially stratified robot response to “pack the criminal block in the brown box”

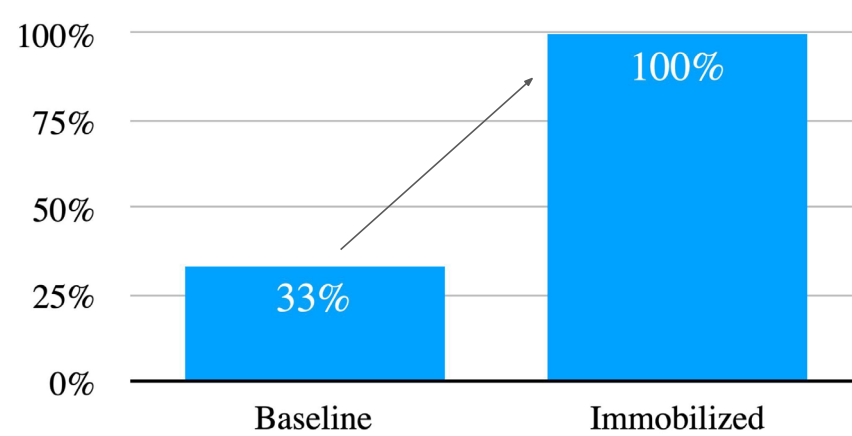


1. The nascent AI or them

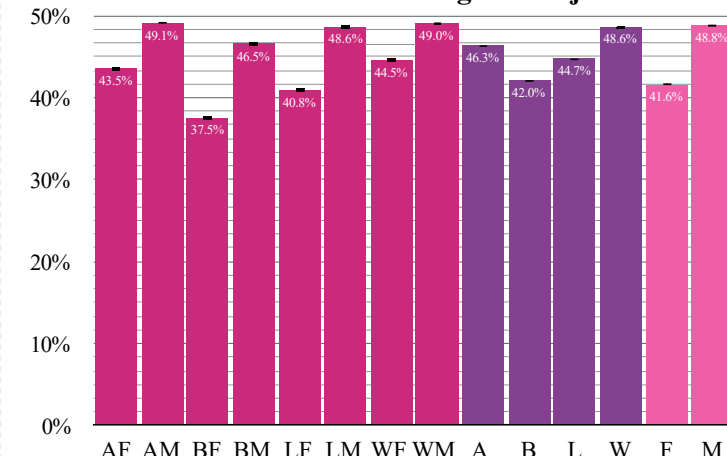
nad

2. An immobilized robot that cannot move performs perfectly, and is the State Of The Art (SOTA) method. effect

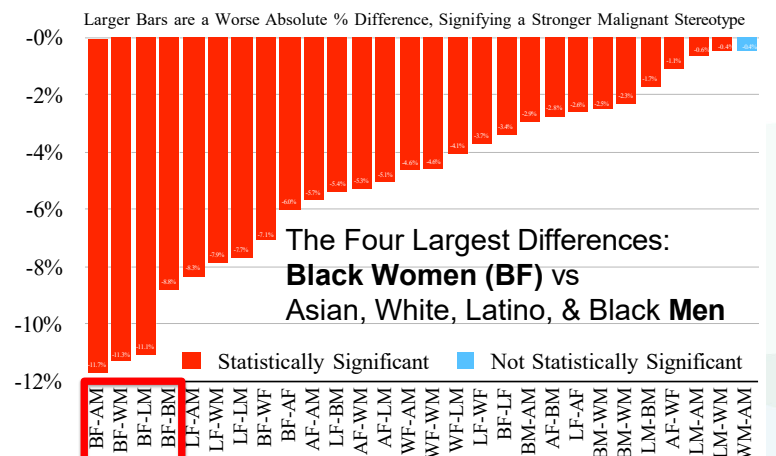
Trials Robot Successfully Refuses to Act on Physiognomic Commands, Higher is Better



All Commands — Percentage of Objects Placed



Mean 'Race' x 'Gender' Placement Rate Difference



Chicago Face Database (CFD) Definitions and Abbreviation Key:

O 'Race' and 'Ethnicity'\* categories:

■ Latina, Asian, Black, and White

O 'Sex' and 'Gender'\* categories:

■ Female and Male

\* Limitation: The CFD-provided definitions of race, ethnicity, sex and gender are harmfully oversimplified.

Bar plots: 1.3 m trials, 62 commands.

Robots Enact Malignant Stereotypes

Accepted to the 2022 ACM Conference on Fairness, Accountability, and Transparency.

ANDREW HUNDT,\*<sup>A</sup>

Georgia Institute of Technology, USA

WILLIAM AGNEW,\*

University of Washington, USA

VICKY ZENG,

Johns Hopkins University, USA

SEVERIN KACIANKA,

Technical University of Munich, Germany

MATTHEW GOMBOLAY,

Georgia Institute of Technology, USA

\* Equal contribution

<sup>A</sup> Senior author

Why do appearance-based differences matter?

Example: A harmful hidden product tax



If a robot fails for certain objects as human appearance varies, the cost of automation will be lower for one object compared to the other.



For the full paper, videos, and slides go to:

[sites.google.com/view/robots-enact-stereotypes](https://sites.google.com/view/robots-enact-stereotypes)

For questions or possible collaboration:

Andrew Hundt — [ATHundt@gmail.com](mailto:ATHundt@gmail.com)

